

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
10 May 2001 (10.05.2001)

PCT

(10) International Publication Number
WO 01/32844 A1

(51) International Patent Classification⁷: **C12N 9/00,**
C12M 1/18

Birkeroed (DK). HUSUM, Tommy, Lykke; Gartnervaaget 18, DK-3400 Hilleroed (DK). OSSOWSKI, Ingemar, Von; University of Helsinki, P.O. Box 56, FIN-00014 Helsinki (FI).

(21) International Application Number: PCT/DK00/00568

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(22) International Filing Date: 10 October 2000 (10.10.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
PA 1999 01605 5 November 1999 (05.11.1999) DK

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(71) Applicant: NOVOZYMES A/S [DK/DK]; Krogshoejvej 36, DK-2880 Bagsværd (DK).

(72) Inventors: PEDERSEN, Henrik; Frodesvej 24, DK-2880 Bagsværd (DK). LAMSA, Michael; 641 Bluebird Place, Davis, CA 95616 (US). HANSEN, Peter, Kamp; Store Stensager 22, DK-4320 Lejre (DK). FRISNER, Henrik; Kurlandsgade 26, 3. th., DK-2300 Copenhagen S (DK). VIND, Jesper; Bagsværdvej 115, DK-2800 Lyngby (DK). ERNST, Steffen; Jesper Brochmands Gade 6, 1. th, DK-2200 Copenhagen N (DK). KONGSBAK, Lars; Vaengestien 2A, DK-2840 Holte (DK). JOERGENSEN, Birthe, Ravn; Fruevej 18, DK-2860 Soeborg (DK). BECK, Thomas, Christian; Bakkevej 11E, DK-3460

Published:

- With international search report.
- Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/32844 A1

(54) Title: MICROTITER PLATE (MTP) BASED HIGH THROUGHPUT SCREENING (HTS) ASSAYS

(57) Abstract: The present invention is to provide a method to perform assays that efficiently and accurately can screen large numbers of cell populations producing variants of a molecule of interest. In a first aspect the invention relates to a method for high throughput screening (HTS) of a large population of host cells for production of a molecule of interest, the method comprising the steps of: a) arranging the host cells in a spatial array so each position in the spatial array is occupied by one cell, b) cultivating the host cells under growth conditions suitable for HTS, c) assaying each array position for production of the molecule of interest, and d) selecting the cells from those positions where the molecule was produced, as determined in step c).

Microtiter plate (MTP) based High Throughput Screening (HTS) assays

Field of invention

5 Technologies such as DNA shuffling, random DNA mutagenesis and in vivo recombination have allowed the generation of enormous populations of variant cells that produce variants of a certain protein, RNA, or small molecule. In addition, it has become possible in recombinant host strains to establish large 10 libraries of natural enzymes cloned from other organisms. Together these technologies have created a need for assays that efficiently and accurately can screen large numbers of variants, high throughput screening (HTS) assays.

15 Background of the invention

Historically, biological assays have been developed and optimized with regard to sensitivity, linearity and simplicity. However, if these assays are to be transformed into high throughput screens for substances produced by large populations 20 cell cultures, a number of difficulties must be overcome. Examples of difficulties encountered when using e.g. a microtiter plate based screening assay include: a) differential growth and/or gene expression of identical clones in the separate wells of the microtiter plate, b) differential growth and/or 25 gene expression of different clones, c) differential stress exposure (eg. heat treatment, humidity) across the plate, d) effects of cell-material or growth medium on the assay. In the past, efficient and accurate high throughput screening has been hampered by the inability to work around such difficulties.

30 Below we describe a method of performing an assay that has been specifically developed for the screening of large populations of clones producing variants of a given protein.

Summary of the invention

The problem to be solved by the present invention is to provide a method to perform assays that efficiently and accurately can screen large numbers of cell populations producing variants of a molecule of interest.

5 In a first aspect the invention relates to a method for high throughput screening (HTS) of a large population of host cells for production of a molecule of interest, the method comprising the steps of:

- 10 a) arranging the host cells in a spatial array so each position in the spatial array is occupied by one cell,
- b) cultivating the host cells under growth conditions suitable for HTS,
- c) assaying each array position for production of the molecule of interest, and
- 15 d) selecting the cells from those positions where the molecule was produced, as determined in step c).

In a second aspect the invention relates to a method for screening a DNA library for DNA of interest, the method comprising the steps of:

- a) creating host cells comprising the DNA library,
- b) using the method described in the first aspect to assay for a host cell producing a molecule of interest; wherein the host cell comprises the DNA of interest.

25

Further, in a third aspect the invention relates to a recombinant vector comprising DNA isolated by a method as defined in the preceding aspect.

In a fourth aspect the invention also relates to a recombinant host cell comprising DNA isolated by a method as described in the second aspect or the vector according to the third aspect.

Also a fifth aspect relates to a transgenic animal containing and expressing the DNA of interest according to any of the 35 second, third or fourth aspect.

A sixth aspect relates to a transgenic plant containing and expressing the DNA of interest according to any of the second, third or fourth aspect.

A seventh aspect relates to a method of producing a molecule of interest, which method comprises cultivating a cell according to any of the first or fourth aspects in suitable culture medium under conditions permitting expression of the DNA of interest and recovering the resulting molecule from the culture medium.

An eighth aspect relates to a method of producing a molecule of interest, which method comprises recovering the molecule from any part or secrete of/from the transgenic animal according to the fifth aspect.

A final aspect relates to a method of producing a molecule of interest, which method comprises growing a cell of a transgenic plant according to the sixth aspect, and recovering the molecule from the resulting plant.

Definitions

Prior to a discussion of the detailed embodiments of the invention is provided a definition of specific terms related to the main aspects of the invention.

In accordance with the present invention there may be employed conventional molecular biology, microbiology, and recombinant DNA techniques within the skill of the art. Such techniques are explained fully in the literature. See, e.g., Sambrook, Fritsch & Maniatis, *Molecular Cloning: A Laboratory Manual*, Second Edition (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (herein "Sambrook et al., 1989"); *DNA Cloning: A Practical Approach*, Volumes I and II /D.N. Glover ed. 1985); *Oligonucleotide Synthesis* (M.J. Gait ed. 1984); *Nucleic Acid Hybridization* (B.D. Hames & S.J. Higgins eds (1985)); *Transcription And Translation* (B.D. Hames & S.J. Higgins, eds. (1984)); *Animal Cell Culture* (R.I. Freshney, ed.

(1986)); *Immobilized Cells And Enzymes* (IRL Press, (1986)); B. Perbal, *A Practical Guide To Molecular Cloning* (1984).

Isolated: When applied to a protein, the term "isolated" indicates that the protein is found in a condition other than its native environment, such as apart from blood and animal tissue. In a preferred form, the isolated protein is substantially free of other proteins, particularly other proteins of animal origin. It is preferred to provide the proteins in a highly purified form, i.e., greater than 95% pure, more preferably greater than 99% pure. When applied to a polynucleotide molecule, the term "isolated" indicates that the molecule is removed from its natural genetic milieu, and is thus free of other extraneous or unwanted coding sequences, and is in a form suitable for use within genetically engineered protein production systems. Such isolated molecules are those that are separated from their natural environment and include cDNA and genomic clones. Isolated DNA molecules of the present invention are free of other genes with which they are ordinarily associated, and may include naturally occurring 5' and 3' un-translated regions such as promoters and terminators. The identification of associated regions will be evident to one of ordinary skill in the art (see for example, Dynan and Tijan, Nature 316: 774-78, 1985).

25 A "polynucleotide" is a single- or double-stranded polymer of deoxyribonucleotide or ribonucleotide bases read from the 5' to the 3' end. Polynucleotides include RNA and DNA, and may be isolated from natural sources, synthesized *in vitro*, or prepared from a combination of natural and synthetic molecules.

30

A "nucleic acid molecule" refers to the phosphate ester polymeric form of ribonucleosides (adenosine, guanosine, uridine or cytidine; "RNA molecules") or deoxyribonucleosides (deoxyadenosine, deoxyguanosine, deoxythymidine, or deoxycytidine; "DNA molecules") in either single stranded form, or a double-stranded helix. Double stranded DNA-DNA, DNA-RNA and RNA-RNA

helices are possible. The term nucleic acid molecule, and in particular DNA or RNA molecule, refers only to the primary and secondary structure of the molecule, and does not limit it to any particular tertiary or quaternary forms. Thus, this term includes double-stranded DNA found, *inter alia*, in linear or circular DNA molecules (e.g., restriction fragments), plasmids, and chromosomes. In discussing the structure of particular double-stranded DNA molecules, sequences may be described herein according to the normal convention of giving only the sequence 10 in the 5' to 3' direction along the non-transcribed strand of DNA (i.e., the strand having a sequence homologous to the mRNA). A "recombinant DNA molecule" is a DNA molecule that has undergone a molecular biological manipulation.

15 A DNA "coding sequence" is a double-stranded DNA sequence which is transcribed and translated into a polypeptide in a cell *in vitro* or *in vivo* when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a 20 translation stop codon at the 3' (carboxyl) terminus. A coding sequence can include, but is not limited to, prokaryotic sequences, cDNA from eukaryotic mRNA, genomic DNA sequences from eukaryotic (e.g., mammalian) DNA, and even synthetic DNA sequences. If the coding sequence is intended for expression in a 25 eukaryotic cell, a polyadenylation signal and transcription termination sequence will usually be located 3' to the coding sequence.

Expression vector: A DNA molecule, linear or circular, that 30 comprises a segment encoding a polypeptide of interest operably linked to additional segments that provide for its transcription. Such additional segments may include promoter and terminator sequences, and optionally one or more origins of replication, one or more selectable markers, an enhancer, a polyadenylation signal, and the like. Expression vectors are generally 35

derived from plasmid or viral DNA, or may contain elements of both.

Transcriptional and translational control sequences are DNA regulatory sequences, such as promoters, enhancers, terminators, and the like, that provide for the expression of a coding sequence in a host cell. In eukaryotic cells, polyadenylation signals are control sequences.

10 A "secretory signal sequence" is a DNA sequence that encodes a polypeptide (a "secretory peptide" that, as a component of a larger polypeptide, directs the larger polypeptide through a secretory pathway of a cell in which it is synthesized. The larger polypeptide is commonly cleaved to remove the secretory 15 peptide during transit through the secretory pathway.

The term "promoter" is used herein for its art-recognized meaning to denote a portion of a gene containing DNA sequences that provide for the binding of RNA polymerase and initiation of 20 transcription. Promoter sequences are commonly, but not always, found in the 5' non-coding regions of genes.

"Operably linked", when referring to DNA segments, indicates that the segments are arranged so that they function in concert 25 for their intended purposes, e.g. transcription initiates in the promoter and proceeds through the coding segment to the terminator.

A coding sequence is "under the control" of transcriptional and 30 translational control sequences in a cell when RNA polymerase transcribes the coding sequence into mRNA, which is then trans-RNA spliced and translated into the protein encoded by the coding sequence.

35 "Isolated polypeptide": is a polypeptide which is essentially free of other non-[enzyme] polypeptides, e.g., at least about

20% pure, preferably at least about 40% pure, more preferably about 60% pure, even more preferably about 80% pure, most preferably about 90% pure, and even most preferably about 95% pure, as determined by SDS-PAGE.

5

"Heterologous" DNA refers to DNA not naturally located in the cell, or in a chromosomal site of the cell. Preferably, the heterologous DNA includes a gene foreign to the cell.

10 A cell has been "transfected" by exogenous or heterologous DNA when such DNA has been introduced inside the cell. A cell has been "transformed" by exogenous or heterologous DNA when the transfected DNA effects a phenotypic change. Preferably, the transforming DNA should be integrated (covalently linked) into
15 chromosomal DNA making up the genome of the cell.

A "clone" is a population of cells derived from a single cell or common ancestor by mitosis.

20 "Homologous recombination" refers to the insertion of a foreign DNA sequence of a vector in a chromosome. Preferably, the vector targets a specific chromosomal site for homologous recombination. For specific homologous recombination, the vector will contain sufficiently long regions of homology to
25 sequences of the chromosome to allow complementary binding and incorporation of the vector into the chromosome. Longer regions of homology, and greater degrees of sequence similarity, may increase the efficiency of homologous recombination.

30 Nucleic Acid Sequence

The techniques used to isolate or clone a nucleic acid sequence encoding a polypeptide are known in the art and include isolation from genomic DNA, preparation from cDNA, or a combination thereof. The cloning of the nucleic acid sequences
35 of the present invention from such genomic DNA can be effected, e.g., by using the well-known polymerase chain reaction (PCR)

or antibody screening of expression libraries to detect cloned DNA fragments with shared structural features. See, e.g., Innis et al., 1990, *A Guide to Methods and Application*, Academic Press, New York. Other nucleic acid amplification procedures such as ligase chain reaction (LCR), ligated activated transcription (LAT) and nucleic acid sequence-based amplification (NASBA) may be used. The nucleic acid sequence may be cloned from a strain producing the polypeptide, or from another related organism and thus, for example, may be an allelic or species variant of the polypeptide encoding region of the nucleic acid sequence.

The term "isolated" nucleic acid sequence as used herein refers to a nucleic acid sequence which is essentially free of other nucleic acid sequences, e.g., at least about 20% pure, preferably at least about 40% pure, more preferably about 60% pure, even more preferably about 80% pure, most preferably about 90% pure, and even most preferably about 95% pure, as determined by agarose gel electrophoresis. For example, an isolated nucleic acid sequence can be obtained by standard cloning procedures used in genetic engineering to relocate the nucleic acid sequence from its natural location to a different site where it will be reproduced. The cloning procedures may involve excision and isolation of a desired nucleic acid fragment comprising the nucleic acid sequence encoding the polypeptide, insertion of the fragment into a vector molecule, and incorporation of the recombinant vector into a host cell where multiple copies or clones of the nucleic acid sequence will be replicated. The nucleic acid sequence may be of genomic, cDNA, RNA, semi-synthetic, synthetic origin, or any combinations thereof.

30

Nucleic Acid Construct

As used herein the term "nucleic acid construct" is intended to indicate any nucleic acid molecule of cDNA, genomic DNA, synthetic DNA or RNA origin. The term "construct" is in-

tended to indicate a nucleic acid segment which may be single- or double-stranded, and which may be based on a complete or partial naturally occurring nucleotide sequence encoding a polypeptide of interest. The construct may optionally contain 5 other nucleic acid segments.

The DNA of interest may suitably be of genomic or cDNA origin, for instance obtained by preparing a genomic or cDNA library and screening for DNA sequences coding for all or part of the polypeptide by hybridization using synthetic oligonucleotide 10 probes in accordance with standard techniques (cf. Sambrook et al., supra).

The nucleic acid construct may also be prepared synthetically by established standard methods, e.g. the phosphoamidite method described by Beaucage and Caruthers, Tetrahedron Letters 15 22 (1981), 1859 - 1869, or the method described by Matthes et al., EMBO Journal 3 (1984), 801 - 805. According to the phosphoamidite method, oligonucleotides are synthesized, e.g. in an automatic DNA synthesizer, purified, annealed, ligated and cloned in suitable vectors.

20 Furthermore, the nucleic acid construct may be of mixed synthetic and genomic, mixed synthetic and cDNA or mixed genomic and cDNA origin prepared by ligating fragments of synthetic, genomic or cDNA origin (as appropriate), the fragments corresponding to various parts of the entire nucleic acid construct, in accordance with standard techniques.

The nucleic acid construct may also be prepared by polymerase chain reaction using specific primers, for instance as described in US 4,683,202 or Saiki et al., Science 239 (1988), 487 - 491.

30 The term nucleic acid construct may be synonymous with the term expression cassette when the nucleic acid construct contains all the control sequences required for expression of a coding sequence of the present invention. The term "coding sequence" as defined herein is a sequence which is transcribed 35 into mRNA and translated into a polypeptide of the present invention when placed under the control of the above mentioned

control sequences. The boundaries of the coding sequence are generally determined by a translation start codon ATG at the 5'-terminus and a translation stop codon at the 3'-terminus. A coding sequence can include, but is not limited to, DNA, cDNA, 5 and recombinant nucleic acid sequences.

The term "control sequences" is defined herein to include all components which are necessary or advantageous for expression of the coding sequence of the nucleic acid sequence. Each control sequence may be native or foreign to the nucleic acid 10 sequence encoding the polypeptide. Such control sequences include, but are not limited to, a leader, a polyadenylation sequence, a propeptide sequence, a promoter, a signal sequence, and a transcription terminator. At a minimum, the control sequences include a promoter, and transcriptional and translational 15 stop signals. The control sequences may be provided with linkers for the purpose of introducing specific restriction sites facilitating ligation of the control sequences with the coding region of the nucleic acid sequence encoding a polypeptide.

20 The control sequence may be an appropriate promoter sequence, a nucleic acid sequence which is recognized by a host cell for expression of the nucleic acid sequence. The promoter sequence contains transcription and translation control sequences which mediate the expression of the polypeptide. The 25 promoter may be any nucleic acid sequence which shows transcriptional activity in the host cell of choice and may be obtained from genes encoding extracellular or intracellular polypeptides either homologous or heterologous to the host cell.

The control sequence may also be a suitable transcription 30 terminator sequence, a sequence recognized by a host cell to terminate transcription. The terminator sequence is operably linked to the 3' terminus of the nucleic acid sequence encoding the polypeptide. Any terminator which is functional in the host cell of choice may be used in the present invention.

35 The control sequence may also be a polyadenylation sequence, a sequence which is operably linked to the 3' terminus

of the nucleic acid sequence and which, when transcribed, is recognized by the host cell as a signal to add polyadenosine residues to transcribed mRNA. Any polyadenylation sequence which is functional in the host cell of choice may be used in
5 the present invention.

The control sequence may also be a signal peptide-coding region, which codes for an amino acid sequence linked to the amino terminus of the polypeptide which can direct the expressed polypeptide into the cell's secretory pathway of the
10 host cell. The 5' end of the coding sequence of the nucleic acid sequence may inherently contain a signal peptide-coding region naturally linked in translation reading frame with the segment of the coding region which encodes the secreted polypeptide. Alternatively, the 5' end of the coding sequence may
15 contain a signal peptide coding region which is foreign to that portion of the coding sequence which encodes the secreted polypeptide. A foreign signal peptide-coding region may be required where the coding sequence does not normally contain a signal peptide-coding region. Alternatively, the foreign signal pep-
20 tide coding region may simply replace the natural signal peptide coding region in order to obtain enhanced secretion relative to the natural signal peptide coding region normally associated with the coding sequence. The signal peptide-coding re-
gion may be obtained from a glucoamylase or an amylase gene
25 from an *Aspergillus* species, a lipase or proteinase gene from a *Rhizomucor* species, the gene for the alpha-factor from *Sac-
charomyces cerevisiae*, an amylase or a protease gene from a *Ba-
cillus* species, or the calf prothrombin gene. However, any
30 signal peptide coding region capable of directing the expressed polypeptide into the secretory pathway of a host cell of choice may be used in the present invention.

The control sequence may also be a propeptide coding re-
gion, which codes for an amino acid sequence positioned at the amino terminus of a polypeptide. The resultant polypeptide is
35 known as a proenzyme or propolypeptide (or a zymogen in some cases). A propolypeptide is generally inactive and can be con-

verted to mature active polypeptide by catalytic or autocatalytic cleavage of the propeptide from the propolypeptide. The propeptide coding region may be obtained from the *Bacillus subtilis* alkaline protease gene (*aprE*), the *Bacillus subtilis* neutral protease gene (*nprT*), the *Saccharomyces cerevisiae* alpha-factor gene, or the *Myceliophthora thermophilum* laccase gene (WO 95/33836).

The nucleic acid constructs of the present invention may also comprise one or more nucleic acid sequences which encode 10 one or more factors that are advantageous in the expression of the polypeptide, e.g., an activator (e.g., a trans-acting factor), a chaperone, and a processing protease. Any factor that is functional in the host cell of choice may be used in the present invention. The nucleic acids encoding one or more of 15 these factors are not necessarily in tandem with the nucleic acid sequence encoding the polypeptide.

An activator is a protein which activates transcription of a nucleic acid sequence encoding a polypeptide (Kudla et al., 1990, *EMBO Journal* 9:1355-1364; Jarai and Buxton, 1994, *Current Genetics* 26:2238-244; Verdier, 1990, *Yeast* 6:271-297). The nucleic acid sequence encoding an activator may be obtained from the genes encoding *Bacillus stearothermophilus* NprA (*nprA*), *Saccharomyces cerevisiae* heme activator protein 1 (*hap1*), *Saccharomyces cerevisiae* galactose metabolizing protein 25 4 (*gal4*), and *Aspergillus nidulans* ammonia regulation protein (*areA*). For further examples, see Verdier, 1990, *supra* and MacKenzie et al., 1993, *Journal of General Microbiology* 139:2295-2307.

A chaperone is a protein which assists another polypeptide in folding properly (Hartl et al., 1994, *TIBS* 19:20-25; Bergeron et al., 1994, *TIBS* 19:124-128; Demolder et al., 1994, *Journal of Biotechnology* 32:179-189; Craig, 1993, *Science* 260:1902-1903; Gething and Sambrook, 1992, *Nature* 355:33-45; Puig and Gilbert, 1994, *Journal of Biological Chemistry* 269:7764-7771; Wang and Tsou, 1993, *The FASEB Journal* 7:1515-

11157; Robinson et al., 1994, *Bio/Technology* 1:381-384). The nucleic acid sequence encoding a chaperone may be obtained from the genes encoding *Bacillus subtilis* GroE proteins, *Aspergillus oryzae* protein disulphide isomerase, *Saccharomyces cerevisiae* calnexin, *Saccharomyces cerevisiae* BiP/GRP78, and *Saccharomyces cerevisiae* Hsp70. For further examples, see Gething and Sambrook, 1992, *supra*, and Hartl et al., 1994, *supra*.

A processing protease is a protease that cleaves a propeptide to generate a mature biochemically active polypeptide (Enderlin and Ogrydziak, 1994, *Yeast* 10:67-79; Fuller et al., 1989, *Proceedings of the National Academy of Sciences USA* 86:1434-1438; Julius et al., 1984, *Cell* 37:1075-1089; Julius et al., 1983, *Cell* 32:839-852). The nucleic acid sequence encoding a processing protease may be obtained from the genes encoding *Aspergillus niger* Kex2, *Saccharomyces cerevisiae* dipeptidylaminopeptidase, *Saccharomyces cerevisiae* Kex2, and *Yarrowia lipolytica* dibasic processing endoprotease (*xpr6*).

It may also be desirable to add regulatory sequences which allow the regulation of the expression of the polypeptide relative to the growth of the host cell. Examples of regulatory systems are those which cause the expression of the gene to be turned on or off in response to a chemical or physical stimulus, including the presence of a regulatory compound. Regulatory systems in prokaryotic systems would include the *lac*, *tac*, and *trp* operator systems. In yeast, the *ADH2* system or *GAL1* system may be used. In filamentous fungi, the TAKA alpha-amylase promoter, *Aspergillus niger* glucoamylase promoter, and the *Aspergillus oryzae* glucoamylase promoter may be used as regulatory sequences. Other examples of regulatory sequences are those which allow for gene amplification. In eukaryotic systems, these include the dihydrofolate reductase gene which is amplified in the presence of methotrexate, and the metallothionein genes which are amplified with heavy metals. In these cases, the nucleic acid sequence encoding the polypeptide would be placed in tandem with the regulatory sequence.

Promoters

Examples of suitable promoters for directing the transcription of the nucleic acid constructs of the present invention, especially in a bacterial host cell, are the promoters obtained from the *E. coli lac* operon, the *Streptomyces coelicolor* agarase gene (*dagA*), the *Bacillus subtilis* levansucrase gene (*sacB*), the *Bacillus subtilis* alkaline protease gene, the *Bacillus licheniformis* alpha-amylase gene (*amyL*), the *Bacillus stearothermophilus* maltogenic amylase gene (*amyM*), the *Bacillus amyloliquefaciens* alpha-amylase gene (*amyQ*), the *Bacillus amyloliquefaciens* BAN amylase gene, the *Bacillus licheniformis* penicillinase gene (*penP*), the *Bacillus subtilis* *xylA* and *xylB* genes, and the prokaryotic beta-lactamase gene (Villa-Kamaroff et al., 1978, *Proceedings of the National Academy of Sciences USA* 75:3727-3731), as well as the *tac* promoter (DeBoer et al., 1983, *Proceedings of the National Academy of Sciences USA* 80:21-25), or the *Bacillus pumilus* xylosidase gene, or by the phage Lambda *P_R* or *P_L* promoters or the *E. coli lac*, *trp* or *tac* promoters. Further promoters are described in "Useful proteins from recombinant bacteria" in *Scientific American*, 1980, 242:74-94; and in Sambrook et al., 1989, *supra*.

Examples of suitable promoters for directing the transcription of the nucleic acid constructs of the present invention in a filamentous fungal host cell are promoters obtained from the genes encoding *Aspergillus oryzae* TAKA amylase, *Rhizomucor miehei* aspartic proteinase, *Aspergillus niger* neutral alpha-amylase, *Aspergillus niger* acid stable alpha-amylase, *Aspergillus niger* or *Aspergillus awamori* glucoamylase (*glaA*), *Rhizomucor miehei* lipase, *Aspergillus oryzae* alkaline protease, *Aspergillus oryzae* triose phosphate isomerase, *Aspergillus nidulans* acetamidase, *Fusarium oxysporum* trypsin-like protease (as described in U.S. Patent No. 4,288,627, which is incorporated herein by reference), and hybrids thereof. Particularly preferred promoters for use in filamentous fungal host cells

are the TAKA amylase, NA2-tpi (a hybrid of the promoters from the genes encoding *Aspergillus niger* neutral α -amylase and *Aspergillus oryzae* triose phosphate isomerase), and *glaA* promoters. Further suitable promoters for use in filamentous fungus host cells are the ADH3 promoter (McKnight et al., The EMBO J. 4 (1985), 2093 - 2099) or the tpiA promoter.

Examples of suitable promoters for use in yeast host cells include promoters from yeast glycolytic genes (Hitzeman et al., J. Biol. Chem. 255 (1980), 12073 - 12080; Alber and Kawasaki, J. Mol. Appl. Gen. 1 (1982), 419 - 434) or alcohol dehydrogenase genes (Young et al., in Genetic Engineering of Microorganisms for Chemicals (Hollaender et al, eds.), Plenum Press, New York, 1982), or the TPI1 (US 4,599,311) or ADH2-4c (Russell et al., Nature 304 (1983), 652 - 654) promoters.

Further useful promoters are obtained from the *Saccharomyces cerevisiae* enolase (ENO-1) gene, the *Saccharomyces cerevisiae* galactokinase gene (GAL1), the *Saccharomyces cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase genes (ADH2/GAP), and the *Saccharomyces cerevisiae* 3-phosphoglycerate kinase gene. Other useful promoters for yeast host cells are described by Romanos et al., 1992, Yeast 8:423-488. In a mammalian host cell, useful promoters include viral promoters such as those from Simian Virus 40 (SV40), Rous sarcoma virus (RSV), adenovirus, and bovine papilloma virus (BPV).

Examples of suitable promoters for directing the transcription of the DNA encoding the polypeptide of the invention in mammalian cells are the SV40 promoter (Subramani et al., Mol. Cell Biol. 1 (1981), 854 -864), the MT-1 (metallothionein gene) promoter (Palmiter et al., Science 222 (1983), 809 - 814) or the adenovirus 2 major late promoter.

An example of a suitable promoter for use in insect cells is the polyhedrin promoter (US 4,745,051; Vasuvedan et al., FEBS Lett. 311, (1992) 7 - 11), the P10 promoter (J.M. Vlak et al., J. Gen. Virology 69, 1988, pp. 765-776), the *Autographa californica* polyhedrosis virus basic protein promoter (EP 397

485), the baculovirus immediate early gene 1 promoter (US 5,155,037; US 5,162,222), or the baculovirus 39K delayed-early gene promoter (US 5,155,037; US 5,162,222).

5 Terminators

Preferred terminators for filamentous fungal host cells are obtained from the genes encoding *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* glucoamylase, *Aspergillus nidulans* anthranilate synthase, *Aspergillus niger* alpha-glucosidase, and 10 *Fusarium oxysporum* trypsin-like protease for fungal hosts) the TPI1 (Alber and Kawasaki, op. cit.) or ADH3 (McKnight et al., op. cit.) terminators.

Preferred terminators for yeast host cells are obtained from the genes encoding *Saccharomyces cerevisiae* enolase, *Sac-15 charomyces cerevisiae* cytochrome C (CYC1), or *Saccharomyces cerevisiae* glyceraldehyde-3-phosphate dehydrogenase. Other useful terminators for yeast host cells are described by Romanos et al., 1992, *supra*.

20 Polyadenylation Signals

Preferred polyadenylation sequences for filamentous fungal host cells are obtained from the genes encoding *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* glucoamylase, *Aspergil-25 lus nidulans* anthranilate synthase, and *Aspergillus niger* al-pha-glucosidase.

Useful polyadenylation sequences for yeast host cells are described by Guo and Sherman, 1995, *Molecular Cellular Biology* 15:5983-5990.

Polyadenylation sequences are well known in the art for 30 mammalian host cells such as SV40 or the adenovirus 5 Elb re-gion.

Signal Sequences

An effective signal peptide-coding region for bacterial 35 host cells is the signal peptide-coding region obtained from the maltogenic amylase gene from *Bacillus* NCIB 11837, the Ba-

cillus stearothermophilus alpha-amylase gene, the *Bacillus licheniformis* subtilisin gene, the *Bacillus licheniformis* beta-lactamase gene, the *Bacillus stearothermophilus* neutral proteases genes (*nprT*, *nprS*, *nprM*), and the *Bacillus subtilis* PrsA gene. Further signal peptides are described by Simonen and Palva, 1993, *Microbiological Reviews* 57: 109-137.

An effective signal peptide coding region for filamentous fungal host cells is the signal peptide coding region obtained from *Aspergillus oryzae* TAKA amylase gene, *Aspergillus niger* neutral amylase gene, the *Rhizomucor miehei* aspartic proteinase gene, the *Humicola lanuginosa* cellulase or lipase gene, or the *Rhizomucor miehei* lipase or protease gene, *Aspergillus* sp. amylase or glucoamylase, a gene encoding a *Rhizomucor miehei* lipase or protease. The signal peptide is preferably derived from a gene encoding *A. oryzae* TAKA amylase, *A. niger* neutral α -amylase, *A. niger* acid-stable amylase, or *A. niger* glucoamylase.

Useful signal peptides for yeast host cells are obtained from the genes for *Saccharomyces cerevisiae* a-factor and *Saccharomyces cerevisiae* invertase. Other useful signal peptide coding regions are described by Romanos et al., 1992, *supra*.

For secretion from yeast cells, the secretory signal sequence may encode any signal peptide which ensures efficient direction of the expressed polypeptide into the secretory pathway of the cell. The signal peptide may be naturally occurring signal peptide, or a functional part thereof, or it may be a synthetic peptide. Suitable signal peptides have been found to be the a-factor signal peptide, the signal peptide of mouse salivary amylase (cf. O. Hagenbuchle et al., *Nature* 289, 1981, pp. 643-646), a modified carboxypeptidase signal peptide (cf. L.A. Valls et al., *Cell* 48, 1987, pp. 887-897), the yeast BAR1 signal peptide (cf. WO 87/02670), or the yeast aspartic protease 3 (YAP3) signal peptide (cf. M. Egel-Mitani et al., *Yeast* 6, 1990, pp. 127-137).

For efficient secretion in yeast, a sequence encoding a leader peptide may also be inserted downstream of the signal sequence and upstream of the DNA sequence encoding the polypeptide. The function of the leader peptide is to allow the expressed polypeptide to be directed from the endoplasmic reticulum to the Golgi apparatus and further to a secretory vesicle for secretion into the culture medium (i.e. exportation of the polypeptide across the cell wall or at least through the cellular membrane into the periplasmic space of the yeast cell). The leader peptide may be the yeast a-factor leader (the use of which is described in e.g. US 4,546,082, EP 123 294, EP 123 544 and EP 163 529). Alternatively, the leader peptide may be a synthetic leader peptide, which is to say a leader peptide not found in nature. Synthetic leader peptides may, for instance, be constructed as described in WO 89/02463 or WO 92/11378.

For use in insect cells, the signal peptide may conveniently be derived from an insect gene (cf. WO 90/05783), such as the lepidopteran *Manduca sexta* adipokinetic hormone precursor signal peptide (cf. US 5,023,328).

20

Expression Vectors

The present invention also relates to recombinant expression vectors comprising a nucleic acid sequence of the present invention, a promoter, and transcriptional and translational stop signals. The various nucleic acid and control sequences described above may be joined together to produce a recombinant expression vector which may include one or more convenient restriction sites to allow for insertion or substitution of the nucleic acid sequence encoding the polypeptide at such sites. Alternatively, the nucleic acid sequence of the present invention may be expressed by inserting the nucleic acid sequence or a nucleic acid construct comprising the sequence into an appropriate vector for expression. In creating the expression vector, the coding sequence is located in the vector so that the

coding sequence is operably linked with the appropriate control sequences for expression, and possibly secretion.

The recombinant expression vector may be any vector (e.g., a plasmid or virus) which can be conveniently subjected 5 to recombinant DNA procedures and can bring about the expression of the nucleic acid sequence. The choice of the vector will typically depend on the compatibility of the vector with the host cell into which the vector is to be introduced. The vectors may be linear or closed circular plasmids. The vector 10 may be an autonomously replicating vector, i.e., a vector which exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, e.g., a plasmid, an extrachromosomal element, a mini-chromosome, or an artificial chromosome. The vector may contain any means for assuring self- 15 replication. Alternatively, the vector may be one which, when introduced into the host cell, is integrated into the genome and replicated together with the chromosome(s) into which it has been integrated. The vector system may be a single vector or plasmid or two or more vectors or plasmids which together 20 contain the total DNA to be introduced into the genome of the host cell, or a transposon.

The vectors of the present invention preferably contain one or more selectable markers which permit easy selection of transformed cells. A selectable marker is a gene the product of 25 which provides for biocide or viral resistance, resistance to heavy metals, prototrophy to auxotrophs, and the like. Examples of bacterial selectable markers are the *dal* genes from *Bacillus subtilis* or *Bacillus licheniformis*, or markers which confer antibiotic resistance such as ampicillin, kanamycin, chloram- 30 phenicol, tetracycline, neomycin, hygromycin or methotrexate resistance. A frequently used mammalian marker is the dihydro-folate reductase gene (DHFR). Suitable markers for yeast host cells are *ADE2*, *HIS3*, *LEU2*, *LYS2*, *MET3*, *TRP1*, and *URA3*. A selectable marker for use in a filamentous fungal host cell may 35 be selected from the group including, but not limited to, *amdS* (acetamidase), *argB* (ornithine carbamoyltransferase), *bar*

(phosphinothricin acetyltransferase), *hygB* (hygromycin phosphotransferase), *niaD* (nitrate reductase), *pyrG* (orotidine-5'-phosphate decarboxylase), *sC* (sulfate adenyltransferase), *trpC* (anthranilate synthase), and glufosinate resistance markers, as well as equivalents from other species. Preferred for use in an *Aspergillus* cell are the *amdS* and *pyrG* markers of *Aspergillus nidulans* or *Aspergillus oryzae* and the *bar* marker of *Streptomyces hygroscopicus*. Furthermore, selection may be accomplished by co-transformation, e.g., as described in WO 91/17243, where the selectable marker is on a separate vector.

The vectors of the present invention preferably contain an element(s) that permits stable integration of the vector into the host cell genome or autonomous replication of the vector in the cell independent of the genome of the cell.

The vectors of the present invention may be integrated into the host cell genome when introduced into a host cell. For integration, the vector may rely on the nucleic acid sequence encoding the polypeptide or any other element of the vector for stable integration of the vector into the genome by homologous or non-homologous recombination. Alternatively, the vector may contain additional nucleic acid sequences for directing integration by homologous recombination into the genome of the host cell. The additional nucleic acid sequences enable the vector to be integrated into the host cell genome at a precise location(s) in the chromosome(s). To increase the likelihood of integration at a precise location, the integrational elements should preferably contain a sufficient number of nucleic acids, such as 100 to 1,500 base pairs, preferably 400 to 1,500 base pairs, and most preferably 800 to 1,500 base pairs, which are highly homologous with the corresponding target sequence to enhance the probability of homologous recombination. The integrational elements may be any sequence that is homologous with the target sequence in the genome of the host cell. Furthermore, the integrational elements may be non-encoding or encoding nucleic acid sequences. On the other hand, the vector may be integrated into the genome of the host cell by non-

homologous recombination. These nucleic acid sequences may be any sequence that is homologous with a target sequence in the genome of the host cell, and, furthermore, may be non-encoding or encoding sequences.

5 For autonomous replication, the vector may further comprise an origin of replication enabling the vector to replicate autonomously in the host cell in question. Examples of bacterial origins of replication are the origins of replication of plasmids pBR322, pUC19, pACYC177, pACYC184, pUB110, pE194, 10 pTA1060, and pAMB1. Examples of origin of replications for use in a yeast host cell are the 2 micron origin of replication, the combination of CEN6 and ARS4, and the combination of CEN3 and ARS1. The origin of replication may be one having a mutation which makes its functioning temperature-sensitive in the 15 host cell (see, e.g., Ehrlich, 1978, *Proceedings of the National Academy of Sciences USA* 75:1433).

More than one copy of a nucleic acid sequence encoding a polypeptide of the present invention may be inserted into the host cell to amplify expression of the nucleic acid sequence. 20 Stable amplification of the nucleic acid sequence can be obtained by integrating at least one additional copy of the sequence into the host cell genome using methods well known in the art and selecting for transformants.

The procedures used to ligate the elements described 25 above to construct the recombinant expression vectors of the present invention are well known to one skilled in the art (see, e.g., Sambrook et al., 1989, *supra*).

Host Cells

30 The present invention also relates to recombinant host cells, comprising a nucleic acid sequence of the invention, which are advantageously used in the recombinant production of the polypeptides. The term "host cell" encompasses any progeny of a parent cell which is not identical to the parent cell due 35 to mutations that occur during replication.

The cell is preferably transformed with a vector comprising a nucleic acid sequence of the invention followed by integration of the vector into the host chromosome. "Transformation" means introducing a vector comprising a nucleic acid sequence of the present invention into a host cell so that the vector is maintained as a chromosomal integrant or as a self-replicating extra-chromosomal vector. Integration is generally considered to be an advantage as the nucleic acid sequence is more likely to be stably maintained in the cell. Integration of the vector into the host chromosome may occur by homologous or non-homologous recombination as described above.

The choice of a host cell will to a large extent depend upon the gene encoding the polypeptide and its source. The host cell may be a unicellular microorganism, e.g., a prokaryote, or a non-unicellular microorganism, e.g., a eukaryote. Useful unicellular cells are bacterial cells such as gram positive bacteria including, but not limited to, a *Bacillus* cell, e.g., *Bacillus alkalophilus*, *Bacillus amyloliquefaciens*, *Bacillus brevis*, *Bacillus circulans*, *Bacillus coagulans*, *Bacillus lautus*, *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus megaterium*, *Bacillus stearothermophilus*, *Bacillus subtilis*, and *Bacillus thuringiensis*; or a *Streptomyces* cell, e.g., *Streptomyces lividans* or *Streptomyces murinus*, or gram negative bacteria such as *E. coli* and *Pseudomonas* sp. In a preferred embodiment, the bacterial host cell is a *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus stearothermophilus* or *Bacillus subtilis* cell. The transformation of a bacterial host cell may, for instance, be effected by protoplast transformation (see, e.g., Chang and Cohen, 1979, *Molecular General Genetics* 168:111-115), by using competent cells (see, e.g., Young and Spizizin, 1961, *Journal of Bacteriology* 81:823-829, or Dubnar and Davidoff-Abelson, 1971, *Journal of Molecular Biology* 56:209-221), by electroporation (see, e.g., Shigekawa and Dower, 1988, *Biotechniques* 6:742-751), or by conjugation (see, e.g., Koehler and Thorne, 1987, *Journal of Bacteriology* 169:5771-5278).

The host cell may be a eukaryote, such as a mammalian cell, an insect cell, a plant cell or a fungal cell.

Useful mammalian cells include Chinese hamster ovary (CHO) cells, HeLa cells, baby hamster kidney (BHK) cells, COS 5 cells, or any number of other immortalized cell lines available, e.g., from the American Type Culture Collection.

Examples of suitable mammalian cell lines are the COS (ATCC CRL 1650 and 1651), BHK (ATCC CRL 1632, 10314 and 1573, ATCC CCL 10), CHL (ATCC CCL39) or CHO (ATCC CCL 61) cell lines. Methods 10 of transfecting mammalian cells and expressing DNA sequences introduced in the cells are described in e.g. Kaufman and Sharp, J. Mol. Biol. 159 (1982), 601 - 621; Southern and Berg, J. Mol. Appl. Genet. 1 (1982), 327 - 341; Loyter et al., Proc. Natl. Acad. Sci. USA 79 (1982), 422 - 426; Wigler et al., Cell 15 14 (1978), 725; Corsaro and Pearson, Somatic Cell Genetics 7 (1981), 603, Ausubel et al., Current Protocols in Molecular Biology, John Wiley and Sons, Inc., N.Y., 1987, Hawley-Nelson et al., Focus 15 (1993), 73; Ciccarone et al., Focus 15 (1993), 80; Graham and van der Eb, Virology 52 (1973), 456; and Neumann 20 et al., EMBO J. 1 (1982), 841 - 845.

In a preferred embodiment, the host cell is a fungal cell. "Fungi" as used herein includes the phyla Ascomycota, Basidiomycota, Chytridiomycota, and Zygomycota (as defined by Hawksworth et al., In, Ainsworth and Bisby's Dictionary of The Fungi, 8th edition, 1995, CAB International, University Press, Cambridge, UK) as well as the Oomycota (as cited in Hawksworth et al., 1995, *supra*, page 171) and all mitosporic fungi (Hawksworth et al., 1995, *supra*). Representative groups of Ascomycota include, e.g., *Neurospora*, *Eupenicillium* (=*Penicillium*), *Emericella* (=*Aspergillus*), *Eurotium* (=*Aspergillus*), and the true yeasts listed above. Examples of Basidiomycota include mushrooms, rusts, and smuts. Representative groups of Chytridiomycota include, e.g., *Allomyces*, *Blastocladiella*, *Ceolomomyces*, and aquatic fungi. Representative groups of Oomycota 35 include, e.g., *Saprolegniomycetous* aquatic fungi (water molds) such as *Achlya*. Examples of mitosporic fungi include *Aspergil-*

lus, *Penicillium*, *Candida*, and *Alternaria*. Representative groups of Zygomycota include, e.g., *Rhizopus* and *Mucor*.

In a preferred embodiment, the fungal host cell is a yeast cell. "Yeast" as used herein includes ascosporogenous yeast (Endomycetales), basidiosporogenous yeast, and yeast belonging to the Fungi Imperfecti (Blastomycetes). The ascosporogenous yeasts are divided into the families Spermophthoraceae and Saccharomycetaceae. The latter is comprised of four sub-families, Schizosaccharomycoideae (e.g., genus *Schizosaccharomyces*), Nadsonioideae, Lipomycoideae, and Saccharomycoideae (e.g., genera *Pichia*, *Kluyveromyces* and *Saccharomyces*). The basidiosporogenous yeasts include the genera *Leucosporidium*, *Rhodosporidium*, *Sporidiobolus*, *Filobasidium*, and *Filobasidiella*. Yeasts belonging to the Fungi Imperfecti are divided into two families, Sporobolomycetaceae (e.g., genera *Sorobolomyces* and *Bullera*) and Cryptococcaceae (e.g., genus *Candida*). Since the classification of yeast may change in the future, for the purposes of this invention, yeast shall be defined as described in *Biology and Activities of Yeast* (Skinner, F.A., Passmore, S.M., and Davenport, R.R., eds, Soc. App. Bacteriol. Symposium Series No. 9, 1980. The biology of yeast and manipulation of yeast genetics are well known in the art (see, e.g., *Biochemistry and Genetics of Yeast*, Bacil, M., Horecker, B.J., and Stopani, A.O.M., editors, 2nd edition, 1987; *The Yeasts*, Rose, A.H., and Harrison, J.S., editors, 2nd edition, 1987; and *The Molecular Biology of the Yeast *Saccharomyces**, Strathern et al., editors, 1981).

The yeast host cell may be selected from a cell of a species of *Candida*, *Kluyveromyces*, *Saccharomyces*, *Schizosaccharomyces*, *Candida*, *Pichia*, *Hansenula*, or *Yarrowia*. In a preferred embodiment, the yeast host cell is a *Saccharomyces carlsbergensis*, *Saccharomyces cerevisiae*, *Saccharomyces diastaticus*, *Saccharomyces douglasii*, *Saccharomyces kluyveri*, *Saccharomyces norbensis* or *Saccharomyces oviformis* cell. Other useful yeast host cells are a *Kluyveromyces lactis* *Kluyveromyces fragilis*

Hansehula polymorpha, *Pichia pastoris* *Yarrowia lipolytica*, *Schizosaccharomyces pombe*, *Ustilgo maylis*, *Candida maltose*, *Pichia guillermondii* and *Pichia methanolic cell* (cf. Gleeson et al., J. Gen. Microbiol. 132, 1986, pp. 3459-3465; US 4,882,279 and US 4,879,231).

In a preferred embodiment, the fungal host cell is a filamentous fungal cell. "Filamentous fungi" include all filamentous forms of the subdivision Eumycota and Oomycota (as defined by Hawksworth et al., 1995, *supra*). The filamentous fungi are characterized by a vegetative mycelium composed of chitin, cellulose, glucan, chitosan, mannan, and other complex polysaccharides. Vegetative growth is by hyphal elongation and carbon catabolism is obligately aerobic. In contrast, vegetative growth by yeasts such as *Saccharomyces cerevisiae* is by budding of a unicellular thallus and carbon catabolism may be fermentative. In a more preferred embodiment, the filamentous fungal host cell is a cell of a species of, but not limited to, *Acremonium*, *Aspergillus*, *Fusarium*, *Humicola*, *Mucor*, *Myceliophthora*, *Neurospora*, *Penicillium*, *Thielavia*, *Tolypocladium*, and *Trichoderma* or a teleomorph or synonym thereof. In an even more preferred embodiment, the filamentous fungal host cell is an *Aspergillus* cell. In another even more preferred embodiment, the filamentous fungal host cell is an *Acremonium* cell. In another even more preferred embodiment, the filamentous fungal host cell is a *Fusarium* cell. In another even more preferred embodiment, the filamentous fungal host cell is a *Humicola* cell. In another even more preferred embodiment, the filamentous fungal host cell is a *Mucor* cell. In another even more preferred embodiment, the filamentous fungal host cell is a *Myceliophthora* cell. In another even more preferred embodiment, the filamentous fungal host cell is a *Neurospora* cell. In another even more preferred embodiment, the filamentous fungal host cell is a *Penicillium* cell. In another even more preferred embodiment, the filamentous fungal host cell is a *Thielavia* cell. In another even more preferred embodiment, the filamentous fungal host cell is a *Thielavia* cell. In another even more preferred embodiment, the filamentous fungal host cell is a *Thielavia* cell.

host cell is a *Tolypocladium* cell. In another even more preferred embodiment, the filamentous fungal host cell is a *Trichoderma* cell. In a most preferred embodiment, the filamentous fungal host cell is an *Aspergillus awamori*, *Aspergillus foetidus*, *Aspergillus japonicus*, *Aspergillus niger*, *Aspergillus nidulans* or *Aspergillus oryzae* cell. In another most preferred embodiment, the filamentous fungal host cell is a *Fusarium* cell of the section *Discolor* (also known as the section *Fusarium*). For example, the filamentous fungal parent cell may be a *Fusarium bactridiooides*, *Fusarium cerealis*, *Fusarium crookwellense*, *Fusarium culmorum*, *Fusarium graminearum*, *Fusarium graminum*, *Fusarium heterosporum*, *Fusarium negundi*, *Fusarium reticulatum*, *Fusarium roseum*, *Fusarium sambucinum*, *Fusarium sarcochroum*, *Fusarium sulphureum*, or *Fusarium trichotheciooides* cell. In another preferred embodiment, the filamentous fungal parent cell is a *Fusarium* strain of the section *Elegans*, e.g., *Fusarium oxysporum*. In another most preferred embodiment, the filamentous fungal host cell is a *Humicola insolens* or *Humicola lanuginosa* cell. In another most preferred embodiment, the filamentous fungal host cell is a *Mucor miehei* cell. In another most preferred embodiment, the filamentous fungal host cell is a *Myceliophthora thermophilum* cell. In another most preferred embodiment, the filamentous fungal host cell is a *Neurospora crassa* cell. In another most preferred embodiment, the filamentous fungal host cell is a *Penicillium purpurogenum* cell. In another most preferred embodiment, the filamentous fungal host cell is a *Thielavia terrestris* cell or an *Acremonium chrysogenum* cell. In another most preferred embodiment, the *Trichoderma* cell is a *Trichoderma harzianum*, *Trichoderma koningii*, *Trichoderma longibrachiatum*, *Trichoderma reesei* or *Trichoderma viride* cell. The use of *Aspergillus* spp. for the expression of proteins is described in, e.g., EP 272 277, EP 230 023.

35 Transformation

Fungal cells may be transformed by a process involving protoplast formation, transformation of the protoplasts, and regeneration of the cell wall in a manner known *per se*. Suitable procedures for transformation of *Aspergillus* host cells 5 are described in EP 238 023 and Yelton et al., 1984, *Proceedings of the National Academy of Sciences USA* 81:1470-1474. A suitable method of transforming *Fusarium* species is described by Malardier et al., 1989, *Gene* 78:147-156 or in co-pending US Serial No. 08/269,449. Examples of other fungal cells are cells 10 of filamentous fungi, e.g. *Aspergillus* spp., *Neurospora* spp., *Fusarium* spp. or *Trichoderma* spp., in particular strains of *A. oryzae*, *A. nidulans* or *A. niger*. The use of *Aspergillus* spp. for the expression of proteins is described in, e.g., EP 272277 or EP 230023. The transformation of *F. oxysporum* may for instance 15 be carried out as described by Malardier et al., 1989, *Gene* 78: 147-156. The transformation of *Trichoderma* spp. may be performed as known in the art.

Yeast may be transformed using the procedures described by Becker and Guarente, In Abelson, J.N. and Simon, M.I., editors, *Guide to Yeast Genetics and Molecular Biology*, Methods in Enzymology, Volume 194, pp 182-187, Academic Press, Inc., New York; Ito et al., 1983, *Journal of Bacteriology* 153:163; and Hinnen et al., 1978, *Proceedings of the National Academy of Sciences USA* 75:1920. Mammalian cells may be transformed by direct uptake using the calcium phosphate precipitation method of Graham and Van der Eb (1978, *Virology* 52:546).

Transformation of insect cells and production of heterologous polypeptides therein may be performed as described in US 4,745,051; US 4,775,624; US 4,879,236; US 5,155,037; US 30 5,162,222; EP 397,485, all of which are incorporated herein by reference. The insect cell line used as the host may suitably be a Lepidoptera cell line, such as *Spodoptera frugiperda* cells or *Trichoplusia ni* cells (cf. US 5,077,214). Culture conditions may suitably be as described in, for instance, WO 89/01029 or 35 WO 89/01028, or any of the aforementioned references.

Transgenic animals

It is also within the scope of the present invention to employ transgenic animal technology to produce the present polypeptide. A transgenic animal is one in whose genome a heterologous DNA sequence has been introduced. In particular, the polypeptide of the invention may be expressed in the mammary glands of a non-human female mammal, in particular one which is known to produce large quantities of milk. Examples of preferred mammals are livestock animals such as goats, sheep and cattle, although smaller mammals such as mice, rabbits or rats may also be employed.

The DNA sequence encoding the present polypeptide may be introduced into the animal by any one of the methods previously described for the purpose. For instance, to obtain expression in a mammary gland, a transcription promoter from a milk protein gene is used. Milk protein genes include the genes encoding casein (cf. US 5,304,489), beta-lactoglobulin, alpha-lactalbumin and whey acidic protein. The currently preferred promoter is the beta-lactoglobulin promoter (cf. Whitelaw et al., Biochem J. 286, 1992, pp. 31-39).

It is generally recognized in the art that DNA sequences lacking introns are poorly expressed in transgenic animals in comparison with those containing introns (cf. Brinster et al., Proc. Natl. Acad. Sci. USA 85, 1988, pp. 836-840; Palmiter et al., Proc. Natl. Acad. Sci. USA 88, 1991, pp. 478-482; Whitelaw et al., Transgenic Res. 1, 1991, pp. 3-13; WO 89/01343; WO 91/02318). For expression in transgenic animals, it is therefore preferred, whenever possible, to use genomic sequences containing all or some of the native introns of the gene encoding the polypeptide of interest. It may also be preferred to include at least some introns from, e.g. the beta-lactoglobulin gene. One such region is a DNA segment which provides for intron splicing and RNA polyadenylation from the 3' non-coding region of the ovine beta-lactoglobulin gene. When substituted for the native 3' non-coding sequences of a gene, this segment

may will enhance and stabilise expression levels of the polypeptide of interest. It may also be possible to replace the region surrounding the initiation codon of the polypeptide of interest with corresponding sequences of a milk protein gene.
5 Such replacement provides a putative tissue-specific initiation environment to enhance expression.

For expression of the present polypeptide in transgenic animals, a nucleotide sequence encoding the polypeptide is operably linked to additional DNA sequences required for its expression to produce expression units. Such additional sequences include a promoter as indicated above, as well as sequences providing for termination of transcription and polyadenylation of mRNA. The expression unit further includes a DNA sequence encoding a secretory signal sequence operably linked to the sequence encoding the polypeptide. The secretory signal sequence may be one native to the polypeptide or may be that of another protein such as a milk protein (cf. von Heijne et al., Nucl. Acids Res. 14, 1986, pp. 4683-4690; and US 4,873,316).

Construction of the expression unit for use in transgenic animals may conveniently be done by inserting a DNA sequence encoding the present polypeptide into a vector containing the additional DNA sequences, although the expression unit may be constructed by essentially any sequence of ligations. It is particularly convenient to provide a vector containing a DNA sequence encoding a milk protein and to replace the coding region for the milk protein with a DNA sequence coding for the present polypeptide, thereby creating a fusion which includes expression control sequences of the milk protein gene.

The expression unit is then introduced into fertilized ova or early-stage embryos of the selected host species. Introduction of heterologous DNA may be carried out in a number of ways, including microinjection (cf. US 4,873,191), retroviral infection (cf. Jaenisch, Science 240, 1988, pp. 1468-1474) or site-directed integration using embryonic stem cells (reviewed 35 by Bradley et al., Bio/Technology 10, 1992, pp. 534-539). The ova are then implanted into the oviducts or uteri of pseudo-

pregnant females and allowed to develop to term. Offspring carrying the introduced DNA in their germ line can pass the DNA on to their progeny, allowing the development of transgenic herds.

General procedures for producing transgenic animals are known in the art, cf. for instance, Hogan et al., Manipulating the Mouse Embryo: A Laboratory Manual, Cold Spring Harbor Laboratory, 1986; Simons et al., Bio/Technology 6, 1988, pp. 179-183; Wall et al., Biol. Reprod. 32, 1985, pp. 645-651; Buhler et al., Bio/Technology 8, 1990, pp. 140-143; Ebert et al., Bio/Technology 6: 179-183, 1988; Krimpenfort et al., Bio/Technology 9: 844-847, 1991, Wall et al., J. Cell. Biochem. 49: 113-120, 1992; US 4,873,191, US 4,873,316; WO 88/00239, WO 90/05188; WO 92/11757 and GB 87/00458. Techniques for introducing heterologous DNA sequences into mammals and their germ cells were originally developed in the mouse. See, e.g. Gordon et al., Proc. Natl. Acad. Sci. USA 77: 7380-7384, 1980, Gordon and Ruddle, Science 214: 1244-1246, 1981; Palmiter and Brinster, Cell 41: 343-345, 1985; Brinster et al., Proc. Natl. Acad. Sci. USA 82: 4438-4442, 1985; and Hogan et al. (*ibid.*). These techniques were subsequently adapted for use with larger animals, including livestock species (see e.g., WO 88/00239, WO 90/01588 and WO 92/11757; and Simons et al., Bio/Technology 6: 179-183, 1988). To summarize, in the most efficient route used to date in the generation of transgenic mice or livestock, several hundred linear molecules of the DNA of interest are injected into one of the pro-nuclei of a fertilized egg according to techniques which have become standard in the art. Injection of DNA into the cytoplasm of a zygote can also be employed.

30 Transgenic plants

Production in transgenic plants may also be employed. It has previously been described to introduce DNA sequences into plants, which sequences code for protein products imparting to the transformed plants certain desirable properties such as increased resistance against pests, pathogens, herbicides or stress conditions (cf. for instance EP 131 620, EP 205 518, EP

270 355, WO 89/04371 or WO 90/02804), or an improved nutrient value of the plant proteins (cf. for instance EP 90 033, EP 205 518 or WO 89/04371). Furthermore, WO 89/12386 discloses the transformation of plant cells with a gene coding for levansucrase or dextranase, regeneration of the plant (especially a tomato plant) from the cell resulting in fruit products with altered viscosity characteristics.

In the plant cell, the DNA sequence encoding the present polypeptide is under the control of a regulatory sequence which 10 directs the expression of the polypeptide from the DNA sequence in plant cells and intact plants. The regulatory sequence may be either endogenous or heterologous to the host plant cell.

The regulatory sequence may comprise a promoter capable of directing the transcription of the DNA sequence encoding the 15 polypeptide in plants. Examples of promoters which may be used according to the invention are the 35s RNA promoter from cauliflower mosaic virus (CaMV), the class I patatin gene B 33 promoter, the ST-LS1 gene promoter, promoters conferring seed-specific expression, e.g. the phaseolin promoter, or promoters 20 which are activated on wounding, such as the promoter of the proteinase inhibitor II gene or the wun1 or wun2 genes.

The promoter may be operably connected to an enhancer sequence, the purpose of which is to ensure increased transcription of the DNA sequence encoding the polypeptide. Examples of 25 useful enhancer sequences are enhancers from the 5'-upstream region of the 35s RNA of CaMV, the 5'-upstream region of the ST-LS1 gene, the 5'-upstream region of the Cab gene from wheat, the 5'-upstream region of the 1'- and 2'-genes of the TR-DNA of the Ti plasmid pTi ACH5, the 5'-upstream region of the octopine 30 synthase gene, the 5'-upstream region of the leghemoglobin gene, etc.

The regulatory sequence may also comprise a terminator capable of terminating the transcription of the DNA sequence encoding the polypeptide in plants. Examples of suitable terminators are the terminator of the octopine synthase gene of the 35

T-DNA of the Ti-plasmid pTiACH5 of Agrobacterium tumefaciens, of the gene 7 of the T-DNA of the Ti plasmid pTiACH5, of the nopaline synthase gene, of the 35s RNA-coding gene from CaMV or from various plant genes, e.g. the ST-LS1 gene, the Cab genes from wheat, class I and class II patatin genes, etc.

The DNA sequence encoding the polypeptide may also be operably connected to a DNA sequence encoding a leader peptide capable of directing the transport of the expressed polypeptide to a specific cellular compartment (e.g. vacuoles) or to extracellular space. Examples of suitable leader peptides are the leader peptide of proteinase inhibitor II from potato, the leader peptide and an additional about 100 amino acid fragments of patatin, or the transit peptide of various nucleus-encoded proteins directed into chloroplasts (e.g. from the St-LS1 gene, SS-Rubisco genes, etc.) or into mitochondria (e.g. from the ADP/ATP translocator).

Furthermore, the DNA sequence encoding the polypeptide may be modified in the 5' non-translated region resulting in enhanced translation of the sequence. Such modifications may, for instance, result in removal of hairpin loops in RNA of the 5' non-translated region. Translation enhancement may be provided by suitably modifying the omega sequence of tobacco mosaic virus or the leaders of other plant viruses (e.g. BMV, MSV) or of plant genes expressed at high levels (e.g. SS-Rubisco, class I patatin or proteinase inhibitor II genes from potato).

The DNA sequence encoding the polypeptide may furthermore be connected to a second DNA sequence encoding another polypeptide or a fragment thereof in such a way that expression of said DNA sequences results in the production of a fusion protein. When the host cell is a potato plant cell, the second DNA sequence may, for instance, encode patatin or a fragment thereof (such as a fragment of about 100 amino acids).

The plant in which the DNA sequence coding for the polypeptide is introduced may suitably be a dicotyledonous plant, examples of which are a tobacco, potato, tomato, or leguminous

(e.g. bean, pea, soy, alfalfa) plant. It is, however, contemplated that monocotyledonous plants, e.g. cereals, may equally well be transformed with the DNA sequence coding for the enzyme.

5 Procedures for the genetic manipulation of monocotyledonous and dicotyledonous plants are well known. In order to construct foreign genes for their subsequent introduction into higher plants, numerous cloning vectors are available which generally contain a replication system for *E. coli* and a 10 selectable/screenable marker system permitting the recognition of transformed cells. These vectors include e.g. pBR322, the pUC series, pACYC, M13 mp series etc. The foreign sequence may be cloned into appropriate restriction sites. The recombinant plasmid obtained in this way may subsequently be used for the 15 transformation of *E. coli*. Transformed *E. coli* cells may be grown in an appropriate medium, harvested and lysed. The chimeric plasmid may then be re-isolated and analyzed. Analysis of the recombinant plasmid may be performed by e.g. determination of the nucleotide sequence, restriction analysis, electrophoresis and other molecular-biochemical methods. After each manipulation the sequence may be cleaved and ligated to another DNA sequence. Each DNA sequence can be cloned on a separate plasmid DNA. Depending on the way used for transferring the foreign DNA into plant cells other DNA sequences might be of importance. In 20 case the Ti-plasmid or the Ri plasmid of *Agrobacterium tumefaciens* or *Agrobacterium rhizogenes*, at least the right border of the T-DNA may be used, and often both the right and the left borders of the T-DNA of the Ri or Ti plasmid will be present 25 flanking the DNA sequence to be transferred into plant cells.

30 The use of the T-DNA for transferring foreign DNA into plant cells has been described extensively in the prior literature (cf. Gasser and Fraley, 1989, Science 244, 1293 - 1299 and references cited therein). After integration of the foreign DNA into the plant genome, this sequence is fairly stable at the 35 original locus and is usually not lost in subsequent mitotic or meiotic divisions. As a general rule, a selectable marker gene

will be co-transferred in addition to the gene to be transferred, which marker renders the plant cell resistant to certain antibiotics, e.g. kanamycin, hygromycin, G418 etc. This marker permits the recognition of the transformed cells containing the DNA sequence to be transferred compared to non-transformed cells.

Numerous techniques are available for the introduction of DNA into a plant cell. Examples are the *Agrobacterium* mediated transfer, the fusion of protoplasts with liposomes containing the respective DNA, microinjection of foreign DNA, electroporation etc. In case *Agrobacterium* mediated gene transfer is employed, the DNA to be transferred has to be present in special plasmids which are either of the intermediate type or the binary type. Due to the presence of sequences homologous to T-DNA sequences, intermediate vectors may integrate into the Ri- or Ti-plasmid by homologous recombination. The Ri- or Ti-plasmid additionally contains the vir-region which is necessary for the transfer of the foreign gene into plant cells. Intermediate vectors cannot replicate in *Agrobacterium* species and are transferred into *Agrobacterium* by either direct transformation or mobilization by means of helper plasmids (conjugation). (Cf. Gasser and Fraley, op. cit. and references cited therein).

Binary vectors may replicate in both *Agrobacterium* species and *E. coli*. They may contain a selectable marker and a poly-linker region which to the left and right contains the border sequences of the T-DNA of *Agrobacterium rhizogenes* or *Agrobacterium tumefaciens*. Such vectors may be transformed directly into *Agrobacterium* species. The *Agrobacterium* cell serving as the host cell has to contain a vir-region on another plasmid. Additional T-DNA sequences may also be contained in the *Agrobacterium* cell.

The *Agrobacterium* cell containing the DNA sequences to be transferred into plant cells either on a binary vector or in the form of a co-integrate between the intermediate vector and the T-DNA region may then be used for transforming plant cells. Usually either multicellular explants (e.g. leaf discs, stem

segments, roots), single cells (protoplasts) or cell suspensions are co-cultivated with *Agrobacterium* cells containing the DNA sequence to be transferred into plant cells. The plant cells treated with the *Agrobacterium* cells are then selected 5 for the co-transferred resistance marker (e.g. kanamycin) and subsequently regenerated to intact plants. These regenerated plants will then be tested for the presence of the DNA sequences to be transferred.

If the DNA is transferred by e.g. electroporation or mi-
10 croinjection, no special requirements are needed to effect transformation. Simple plasmids e.g. of the pUC series may be used to transform plant cells. Regenerated transgenic plants may be grown normally in a greenhouse or under other conditions. They should display a new phenotype (e.g. production of
15 new proteins) due to the transfer of the foreign gene(s). The transgenic plants may be crossed with other plants which may either be wild-type or transgenic plants transformed with the same or another DNA sequence. Seeds obtained from transgenic plants should be tested to assure that the new genetic trait is
20 inherited in a stable Mendelian fashion.

See also Hiatt, Nature 344: 469-479, 1990; Edelbaum et al., J. Interferon Res. 12: 449-453, 1992; Sijmons et al., Bio/Technology 8: 217-221, 1990; and EP 255 378.

25 Methods of Production

The transformed or transfected host cells described above are cultured in a suitable nutrient medium under conditions permitting the production of the desired molecules, after which these are recovered from the cells, or the culture broth.

30 The medium used to culture the cells may be any conventional medium suitable for growing the host cells, such as minimal or complex media containing appropriate supplements. Suitable media are available from commercial suppliers or may be prepared according to published recipes (e.g. in catalogues 35 of the American Type Culture Collection). The media are pre-

pared using procedures known in the art (see, e.g., references for bacteria and yeast; Bennett, J.W. and LaSure, L., editors, *More Gene Manipulations in Fungi*, Academic Press, CA, 1991).

If the molecules are secreted into the nutrient medium,
5 they can be recovered directly from the medium. If they are not secreted, they can recovered from cell lysates. The molecules are recovered from the culture medium by conventional procedures including separating the host cells from the medium by centrifugation or filtration, precipitating the proteinaceous
10 components of the supernatant or filtrate by means of a salt, e.g. ammonium sulphate, purification by a variety of chromatographic procedures, e.g. ion exchange chromatography, gelfiltration chromatography, affinity chromatography, or the like, dependent on the type of molecule in question.

15 The molecules of interest may be detected using methods known in the art that are specific for the molecules. These detection methods may include use of specific antibodies, formation of a product, or disappearance of a substrate. For example, an enzyme assay may be used to determine the activity of
20 the molecule. Procedures for determining various kinds of activity are known in the art.

The molecules of the present invention may be purified by a variety of procedures known in the art including, but not limited to, chromatography (e.g., ion exchange, affinity, hydrophobic, chromatofocusing, and size exclusion), electrophoretic procedures (e.g., preparative isoelectric focusing (IEF), differential solubility (e.g., ammonium sulfate precipitation), or extraction (see, e.g., *Protein Purification*, J.C. Janson and Lars Ryden, editors, VCH Publishers, New York, 1989).

30

Detailed description of the invention

When cell cultures rather than purified substances are used in assays, a complication arises: Cells, cell particles, growth media or matter secreted by the cells can have an effect on the
35 treatment (e.g., heat treatment) of the substance in question,

on the assay itself or on the detection of the assay signal. The present invention provides solutions to these problems.

As it is not feasible to handle individual clones separately, it also becomes necessary with respect to the assay to level out differences in growth, gene expression and other differing characteristics of individual clones, in order to handle many clones simultaneously in an automated manner.

The present invention solves these problems, by defining conditions of growth so as to achieve a uniform distribution of 10 growth and gene expression in identical clones in the wells of the same microtiter plate (MTP), or by developing a method for doing HTS-assays that are independent of molecule concentrations within a broad range. Solutions to these problems include accurate and simple detection of the concentration of the active molecules, and development of automated assays that incorporate this information in the protocol during the assay, or that produce raw-data that can be corrected by use of this information after the assay has been performed.

Altogether, the solutions provided by this invention 20 makes microtiter plate based high throughput screening of substances produced by cell cultures a practical and reliable method for the identification of novel substances with interesting properties, from large libraries of molecules.

A first preferred embodiment of the invention relates to 25 a method for high throughput screening (HTS) of a large population of host cells for production of a molecule of interest, the method comprising the steps of:

- a) arranging the host cells in a spatial array so each position in the spatial array is occupied by one cell,
- 30 b) cultivating the host cells under growth conditions suitable for HTS,
- c) assaying each array position for production of the molecule of interest, and
- d) selecting the cells from those positions where the molecule was produced, as determined in step c).

A method of the first aspect can be performed in many ways, the spatial array of step a) can take on any physical form whatsoever, that enables the assaying of several samples at once, without one sample contaminating another. Examples of preferred spatial arrays are different kinds of microtiter-plates with any number of wells, such as 96 or 384, and of any kind of material, as well as positions in a High Performance Liquid Chromatography (HPLC) autosampler device. Any kind of 10 physical arrangement which allows the unambiguous identification of the samples by a number or a position in the array. Even samples placed as drops on a surface in a specific recorded pattern, the surface being of a solid material or of more complex nature such as a textile or a tissue, e.g. cotton, 15 wool, paper, or cellulose.

A preferred embodiment relates to a method of the first aspect, wherein the spatial array of point a) is a microtiter plate, a solid surface or a textile surface.

A way of carrying out step c) of the first aspect of the 20 invention could be to take a sample from each position of the spatial array, e.g. from a supernatant or cell, and transfer this to another spatial array for further testing or assaying for production of the molecule of interest (see example 2). The second spatial array may be identical to the first one used in 25 the specific method, but may also be of any other kind that fulfils the above mentioned criteria, such as a microtiter plate, a solid surface, a textile, any material etc.

Accordingly a preferred embodiment relates to a method of the first aspect, wherein after step b) a sample is transferred 30 from each position in the spatial array to a position in a second spatial array which is then used onwards in the method, preferably the second array is a microtiter plate, more preferably a solid surface, and most preferably a textile surface.

It is a difficult task to make protein libraries in filamentous fungi. One problem is that the different morphology and 35 growth rate of different clones give rise to large and small

clones that tend to grow out of their position in the spatial array. Assaying fungal transformants of different sizes is also a difficult task due to differences in amounts of produced molecules. The present invention solves this by growing the individual clones/cells within microenvironments in each position of the spatial array (see example 9). Such microenvironments are initially sterile beads or balls of any material that allows growth of the clone, preferably beads comprising agarose, alginic acid, polysaccharide, carbohydrate, alginate, carrageenan, chitosan, cellulose, pectin, dextran or polyacrylamide, all allowing diffusion to/from each micro-environment.

Accordingly a preferred embodiment relates to a method of the first aspect, wherein each position in the spatial array is occupied by a bead comprising one cell, preferably the bead is an agarose-bead.

The assaying of the first aspect of the invention for production of the molecule of interest can be done in a great number of ways. As indicated, some kinds of spatial arrays like microtiter plates, can sometimes be assayed directly, or samples can be taken from each position and transferred to another spatial array for assaying. The assay as such can be based on a number of techniques well known in the art such as antibody binding, ELISA and many others; common to these assays is that a measurement is taken of a detectable property e.g. fluorescence, luminescence, β -galactosidase or another enzymatic activity etc. The method of the invention can be performed with any number of these assays, and consequently the molecules assayed for in the first aspect step c) can be obtained in a number of ways, depending on the host cell construct. The molecule of interest may be secreted by the host cell into the supernatant, lysis of the host cells may release the molecule or the molecule may remain intracellular.

A preferred embodiment relates to a method of the first aspect, wherein the molecule of interest in the first aspect step c) is assayed in either whole broth, supernatant of cells that se-

crete the molecule, a lysate of cells that produce the molecule, or is assayed while still inside cells that produce the molecule.

As described above, any number of host cells can be used
5 to perform the method of the invention.

Accordingly a preferred embodiment relates to a method of the first aspect, wherein the host cells are of mammalian origin, preferably hybridoma cells.

More preferably the host cells are of microbial origin,
10 preferably bacterial or fungal.

Even more preferably the host cells are chosen from the group consisting of *Escherichia coli*, *Bacillus subtilis*, *Bacillus licheniformis*, *Bacillus clausii*, *Aspergillus niger*, *Aspergillus oryzae*, *Aspergillus nidulans*, and *Saccharomyces cerevisiae*.

The molecule of interest in the first aspect of the invention can be any molecule of biological origin, non-limiting examples of which are peptides, polypeptides, proteins, enzymes, post-translationally modified polypeptides such as lipopeptides or glycosylated peptides, antimicrobial peptides or molecules, primary or secondary metabolites such as alginates, small organic molecules, molecules having pharmaceutical properties etc.

25 Accordingly a preferred embodiment relates to a method of the first aspect, wherein the molecule of interest is chosen from the group consisting of polypeptides, small peptides, lipopeptides, antimicrobials, small molecules and pharmaceutical molecules.

30 Another preferred embodiment relates to a method of the first aspect, wherein the molecule is an enzyme, preferably chosen from the group consisting of amylase, amyloglucosidase, carbohydrase, carboxypeptidase, cellulase, glucoamylase, glucose oxidase, glucosidase, haloperoxidase, hemicellulase, invertase, isomerase, laccase, lipase, lyase, mannosidase, oxi-

dase, pectinolytic enzyme, peroxidase, phytase, phenoloxidase, polyphenoloxidase, protease, and transglutaminase.

An obstacle when working with more than one clone is that different clones produce different levels of different variants, which means that a quantitative comparison of e.g. molecule yield between the clones is difficult. The invention provides a way to overcome this obstacle, by growing the clones under conditions that ensure minimal variation in molecule concentration between the clones. Non-limiting examples of growth conditions that must be kept essentially equal over the entire spatial array, or similar in each position are temperature, component concentrations in solution (equalize the evaporation from each position, see example 1), amount of molecule produced in each position (see example 9) etc.

Accordingly a preferred embodiment relates to a method of the first aspect, wherein the growth conditions of the first aspect step b) ensure minimal variation in the molecule concentration between all positions in the spatial array.

An obstacle to achieve efficient HTS is how to obtain an accurate determination of the amount of specific active molecule produced by each individual clone. One solution is to configure an assay in such a way that it is insensitive to the concentration of the molecule (as exemplified in example 2). Another solution is that when the concentration of the specific molecule in a position of the spatial array has been determined, this information can be used to determine the specific activity from the total activity determined in that position; alternatively, the information can be used to adjust the input-amount of the molecule into the activity assay.

Non-limiting examples of how the concentration of the specific molecule may be determined are: a) Total molecule measurement. In cases where the molecule of interest is secreted and makes up a large proportion of the culture supernatant, a total molecule determination will provide a reasonable estimate for the amount of specific molecule; b) Specific

molecule measurements. The concentration of e.g. a specific molecule such as an enzyme may be determined by any of the following methods (see examples) :

- a. Active site titration of enzymes using a specific enzyme inhibitor.
- b. By using a fluorescently labelled molecule that binds specifically to the molecule of interest, and measure the change in fluorescence polarization as the labelled molecule is added to the culture supernatant containing the molecule of interest,
- c. By ELISA measurements using antibodies against the specific protein, including competitive ELISA,
- d. Co-expression of green fluorescent protein. By transcriptional (or translational) fusion of a detectable protein (e.g., Green Fluorescent Protein, GFP) and measurement of fluorescense in the enzyme/protein solution,
- e. Tagging of the enzyme/protein in question, followed by labelling and subsequent detection (e.g. with fluorescent probe or NMR imaging agent) of the tag.
- f. Alternatively, the amount of enzyme/protein input into the assay may be controlled (kept constant) by the inclusion of a purification step, e.g. Ni-NTA purification of His-tagged proteins or binding of enzyme/protein to specific antibody.

25

Accordingly a preferred embodiment relates to a method of the first aspect, wherein the assay of the first aspect step c) is done under conditions where the assay is essentially independent of the concentration of the molecule.

30 Preferably the molecule is an enzyme, and the enzyme concentration is in a range of preferably 98% below to 400% above the average concentration of enzyme employed in the assay, more preferably 50% below to 200% above the average concentration of enzyme employed in the assay.

Another preferred embodiment relates to a method of the first aspect, wherein the assay done in the first aspect step c) is followed by the additional steps of:

- 5 a) determining the concentration of the molecule of interest for each position; and
- b) adjusting the concentration of molecules in each position to essentially the same level and then performing the assay of the first aspect step c); or
- 10 c) performing the assay in the first aspect step c) and then correcting the data obtained with regard to the concentration of the molecule.

Preferably the molecule of interest is an enzyme, and the concentration in step a) of the preceding embodiment is indirectly determined by a method chosen from the group consisting of active site titration using an enzyme inhibitor, measurement 15 of fluorescence polarization using a fluorescently labelled enzyme inhibitor, measurement of labelled anti-enzyme antibodies, and measurement of a marker protein co-expressed with the enzyme.

20 An additional preferred embodiment relates to a method of the first aspect, wherein the cultivation of the first aspect step b) is followed by an additional step of recovering from each position essentially the same amount of molecule and using these recovered amounts in the assay of the first aspect step 25 c).

Preferably the molecule of the previous embodiment is a polypeptide, and the recovery is achieved by fusing a polyhistidine-tag to the polypeptide and recovering this construct on a Ni-NTA solid support.

30 When a screening method is used, the choice of assay depends on the molecule of interest and on the host cell construct. Many ways are known in the art to assay or monitor the production of a molecule, some *in vivo* others *in vitro*. Usually such assays are based on the determination of a chemical or 35 physical event, which may be directly indicative or indirectly through a cascade of events resulting in e.g. a change of col-

our in the host cell or in the surrounding medium, emission of light, production of detectable secondary molecules like proteins or enzymes, binding of detectable molecules, cell death, cell lysis, cell growth inhibition or promotion (see examples).

5 Non-limiting examples of events are:

1. Cleavage or other modification of a substrate in solution.
2. Cleavage or other modification of a substrate (e.g. a stain), covalently or non-covalently attached to textile or other type of solid material, e.g. ceramics, metal surfaces,
- 10 plastic or other polymer.
3. Cleavage or other modification of substrate cross-linked/immobilized to bacterial cellulose or other type of polymer surface.
4. Cleavage or other modification of non-stained textile
- 15 swatches (e.g., cellulose swatch).
5. Interaction with live cells, resulting in lysis, growth inhibition or growth promotion of these cells.
6. Interaction with another molecule, or group of molecules, leading to a binding event.

20 These events may be detected by e.g. radioactivity, fluorescence, fluorescense polarization, luminiscence, NMR, mass spectrometry, absorbance etc. The substance assayed for may be derived from for example cell culture supernatants, lysed cells or live cells.

25 Accordingly a preferred embodiment relates to a method of the first aspect, wherein the assay used in the first aspect step c) allows the determination of a chemical or physical event in a live cell, preferably lysis, growth inhibition, growth promotion, production of a protein or another molecule

30 by the cell.

Another preferred embodiment relates to a method of the first aspect, wherein the assay used in the first aspect step c) allows the determination of an event in which the molecule of interest interacts with another molecule, preferably leading

to a binding event or a chemical modification of the other molecule.

Yet another preferred embodiment relates to a method of the first aspect, wherein the assay used in the first aspect step c) allows the determination of the cleavage, isomerization, ligation or other modification of a substrate.

Preferably the substrate of the preceding embodiment is a polymer, more preferably chosen from the group consisting of polypeptides, celluloses, polysaccharides, and starches.

10 Also, preferably the substrate of the two preceding embodiments is immobilized to the surface of a solid material, preferably textile or ceramics.

More preferably the substrate is a solid surface, most preferably a cellulose swatch.

15 The substrate of the four preceding embodiments is preferably labelled with a detectable probe.

Preferably the probe of the preceding embodiment is fluorescent, more preferably it is fluorescein-5-isothiocyanate (FITC) or dichlorotriazino-5-aminofluorescein (DTAF).

20 An industrially very interesting area of research, is the optimization of enzymes to be used in washing applications, consequently the method of the invention is particularly interesting in this area of enzyme research. The assay of the first aspect step c) can be carried out under conditions of high pH 25 and elevated temperatures by adding detergent, powder or fluid, to the assay and controlling the temperature as well as the shaking.

Accordingly a preferred embodiment relates to a method of the first aspect, wherein the assay of the first aspect step c) 30 is performed in a solution comprising a detergent and where the molecule of interest is an enzyme.

As describe above, a number of assays well known in the art can be used with the present invention, these could be based on any detectable or measurable property that can be associated with production of the molecule of interest.

Accordingly a preferred embodiment relates to a method of the first aspect, wherein the assay used in the first aspect step c) involves detection of a signal chosen from the group consisting of fluorescence, fluorescence polarization, absorbance, radioactivity, nuclear magnetic resonance (NMR) and luminiscence.

It was mentioned above how varying concentrations of the molecule in the positions of the spatial array could impair the method of the invention, however not only concentration but also total amount can vary, and measures can be taken to equalize total amount produced of the molecule of interest (see examples 4, 5, 7, and 9).

Accordingly a preferred embodiment relates to a method of the first aspect, wherein the growth conditions of the first aspect step b), have been optimized to give minimal variation in the amount produced of the molecule of interest in each position of the spatial array.

A preferred embodiment relates to a method of the first aspect, wherein the growth conditions are chosen so as to mimic the conditions under full-scale fermentations, preferably highly dilute media are chosen.

A method for screening a DNA library for DNA of interest according to the second aspect of the invention. An essential element in this method is a step of using the method of the first aspect as described herein.

In the art, ways of generating and producing DNA libraries from natural sources are well known, but besides natural DNA sequences, a number of ways are also known in which to generate very large populations of diverse artificial DNA sequences starting from one or more natural sequences, e.g. shuffling or directed evolution (WO 98/42832; US 5,965,408; WO 98/01581; WO 97/07205; WO 95/22625; US 5,093,257).

A preferred embodiment relates to a method of the second aspect, wherein the DNA library is generated from a natural DNA sequence by DNA shuffling or directed evolution.

Preferably the host cells of the second aspect step a) are of microbial origin.

A recombinant vector according to the third aspect comprising DNA isolated by a method as defined in the second aspect of the invention.

A recombinant host cell according to the fourth aspect comprising DNA isolated by a method as described in the second aspect or the vector according to the third aspect.

Preferably the cell according to the fourth aspect is of microbial origin, even more preferably of *Bacillus* or *Aspergillus* origin, and most preferably chosen from the group consisting of: *Escherichia coli*, *Bacillus subtilis*, *Bacillus licheniformis*, *Bacillus clausii*, *Aspergillus niger*, *Aspergillus oryzae*, *Aspergillus nidulans*, and *Saccharomyces cerevisiae*.

A transgenic animal according to the fifth aspect containing and expressing the DNA of interest according to the second aspect.

A transgenic plant according to the sixth aspect containing and expressing the DNA of interest according to the second aspect.

The industrial interest in the method of the invention is of course spurred by the potential of producing the molecule of interest according to the aspects of the invention. Hence the invention relates to the ways known in the art of producing such molecules as described in more detail above.

A method of producing a molecule of interest according to the seventh aspect, which method comprises cultivating a cell according to the fourth aspect in suitable culture medium under conditions permitting expression of the DNA of interest and recovering the resulting molecule from the culture medium.

A method of producing a molecule of interest according to the eighth aspect, which method comprises recovering the molecule from any part or secrete of/from the transgenic animal according to the fifth aspect.

A method of producing a molecule of interest according to the final aspect, which method comprises growing a cell of a

transgenic plant according to the sixth aspect, and recovering the molecule from the resulting plant.

Example 1

5 Minimization of evaporation of liquids from microtiter-plates by the use of a culture box.

A culture box was designed in order to minimize evaporation of liquids from miniaturized microtiter plate fermentations and analytical methods at temperatures from 30-45°C, or 10 even 4-99°C. The box is designed with a shallow base to hold water and absorbent paper. A tray is secured to a base that holds stacks of microtiter plates between stainless steel rods. An airtight lid fitted with a silicon gasket is placed over the base/tray and the lid is secured with clamps on all 4 sides. 15 Rubber bands are placed around each set of 4 rods that are used to secure the plates into position, such that when the box is placed in an incubator shaker the entire device moves as a single unit.

Use of the culture box eliminates or minimizes many of 20 the problems associated with miniaturized analytical and fermentation methods in microtiter plates. By limiting evaporation, the growth and expression differences across the plate, particularly the differences between clones in the centre and at the edges are minimized. It allows a more uniform growth in 25 static or shaken cultures, as well as allows adequate oxygenation of cultures by rapid plate shaking (up to 400 RPM) for several days. The use of this box allows for a wide variety of experimental conditions to be performed in microtiter plates, e.g., very low volumes (2 - 20 ml in a 96 well plate) and very 30 dilute solutions (media) may be used without significant loss of volume.

Example 2

Wash performance assay for detergent α -amylase.

Coating of swatches with starch

Unmodified starch from a natural source is mixed with small amounts of fluorescently labelled starch and coated on a solid phase. The natural starch source can be flour derived from e.g. potato, corn, or rice. Especially rice flour has been observed to provide a good correlation to higher scale wash trials. The solid phase may be twill, as twill has been found to provide good correlation to larger scale washes and has good handling properties. The overall starch concentration as well as the ratio between labelled and unlabeled starch can be varied over a wide range, but we have found 5% (w/v) unmodified starch and 0.025% (w/v) fluorescin 5-isothiocyanate (FITC) labelled potato starch (50-300 glucose units per FITC molecule) to provide the best compromise between sensitivity and response level (FITC was obtained from Sigma).

The starches are suspended in water and boiled for 10 min, alternatively jet-cooked for 5 min at 105°C and 2 bar. After cooling, a selected textile is soaked with the cooked suspension, excess slurry removed by rolling, and the textile is dried either overnight at ambient temperature or for a few minutes at high temperature, e.g. 70°C, at high air velocity, e.g. 12 m/s. We have found that automation of this coating procedure to significantly dampen coating variation. By use of a coating machine to continuously coat several hundred meters at a time, subsequently performed assays with repeated dose-response curves exhibit considerably less variation compared to non-continuous coatings.

Growth of *Bacillus* cells secreting α-amylase variants

Bacillus cells secreting α-amylase variants are grown in 2xYT and 6 µg/ml Chloramphenicol for three days at 37°C and 240 rpm, in a culture box (see example 1).

High throughput screen for detergent α-amylases with better wash performance

Dry starch coated textile is punched into wells of a polystyrene microtiter plate, preferentially of the 96-well format. Assays are performed by first applying a detergent solution, e.g. 150 μ l, to each well. A range of detergents can be used such as liquid detergent or powder detergent dissolved in water. Water hardness is controlled by the addition of a desired amount of calcium and magnesium ions. The assay itself is insensitive to water hardness over a wide range, e.g. 0-30°dH. If a detergent that contains enzymes is used, these enzymes can be inactivated by e.g. heating the detergent to 85°C for 5 min prior to the assay. Furthermore, the detergent can be centrifuged and/or filtered before use to minimise particulate matter. However, no assay interference due to un-removed particles has been observed with detergents used in the range of dosages recommended by the manufacturers.

The intrinsic pH of detergents can be enforced with buffering capacity by adding e.g. glycine or 3-[cyclohexylamino]-1-propane-sulfonic acid (CAPS). This is an important quality enhancing measure, since the high throughput screen uses culture supernatants as the source of expressed enzyme. Growth media normally contain buffer components to ensure the optimal growth pH, which is rarely equal to the high pH (often pH 9-11) present in dissolved detergents.

Culture medium from above is added immediately after dispensing the detergent. We prefer to use a small amount of enzyme solution compared to detergent/buffer, in order to minimise artefacts caused by e.g. a rich growth medium. For example, we routinely apply 15 μ l of enzyme solution to 150 μ l detergent.

Following addition of detergent and enzyme (culture medium), the simulated "wash" takes place. Application of mechanic motion such as vigorous shaking (e.g., 240 rpm) at this point strongly promotes the effect of tested enzymes, thus creating a good dynamic range for ranking candidates. Incubation times at this step from 5 min to 2 h have been tested. When screening for amylases at low-temperature conditions, we have

found 15 min "washing" to provide a good window of difference between "good" or "bad" candidates.

Heating can be applied during incubation to simulate the heating of water during machine washing used in many parts of the world. For example, simulation of European washing conditions could include heating up to 40 or 60°C. This heating can be introduced gradually by for example placing the ambient temperature microtiter tray in a shaking incubator set to the appropriate temperature. By this approach, a heat-up profile is generated. Alternatively, if constant high temperature is desired, heated buffers can be added or insulated tubings and thermostated surroundings can be employed.

The assay responses can be read by transferring the wash solution to another microtiter plate and measure released fluorescence in this solution. Alternatively, the wash solution can be completely removed and the response measured as residual fluorescence of the swatch.

By using the conditions described above, the two approaches generate almost exact mirror image data, meaning that a high degree of released fluorescence in a well is reflected by a low degree of residual fluorescence on the corresponding swatch and vice versa. We prefer the last method, as this only requires one well per assay, thus ensuring minimal amounts of operations leading to maximal throughput in e.g. a robotic setup.

The growth conditions used ensure a reasonably uniform growth of identical clones in different wells of the plate. However, large differences in expression level between different variants still exist. We have found that amylase variants with different performance in the final application release different amounts of fluorescent stain from the swatch in the wash performance assay, even when the enzyme is used in saturating concentrations. Therefore, the screen has been set up using very high concentrations of enzyme. Under these conditions the stain removal efficiency depends only minimally on the enzyme concentration within a broad range; consequently,

differences in expression levels between variants have little effect on the performance of the individual variant in the high throughput screening assay.

When applying the assay to libraries of amylase producing clones, hits are defined as wells of the plate wherein the measurements exceed a given assay response relative to a selected benchmark amylase.

Results.

10 Correlation with the more application-relevant mini-wash test.

Orange coloured rice starch textile was shaken vigorously in 60 mL Asia-Pacific model detergent containing an amylase variant, and the reflectance measured. Table 1 below shows the correlation between the performance of three different amylases 15 (1, 2, and 3) in the swatch assay (the High Throughput Screening set-up) and the conventional mini-wash test. The improvement factor in a given assay is given by the performance of the variant amylase divided by the performance of a given benchmark amylase. Obviously, the improved amylase (amylase 3) performs 20 better in both the swatch assay and the mini-wash.

	Amylase 1	Amylase 2	Amylase 3
Swatch assay	1.0	0.4	1.1
Mini-wash	1.0	0.3	1.3

Table 1: Relative residual fluorescence intensity on washed swathes compared to relative reflectance of orange coloured 25 rice starch textile in mini wash. Application dosage of 0.2 µg/ml amylase was used for the mini wash, while approx. 2 µg/ml final dosage was used in the swatch assay (the average concentration from culture supernatants in the High Throughput Assay).

30

Example 3

HTS screening assay for anti-fungal activity

An anti-fungal agent can either inhibit the outgrowth of spores, vegetative cells, or both. To identify an anti-fungal agent inhibiting vegetative cells, 50 µl samples of a liquid culture of a tester strain, e.g. *Botrytis cinerea*, are distributed into each well of a microtiter plate; 50 µl of samples of sterile culture supernatants from cultures of bacterial cells, e.g. *Bacillus*, to be screened for secretion of anti-fungal activities (e.g. variants of an antimicrobial peptide or enzyme), are then added to each well of the microtiter plate. Optionally, instead of sterilizing the culture supernatants, a tissue culture insert (e.g., Nunc TC Insert) may be inserted, to prevent contact between the secreting cells and cells of the tester strain. The insert may contain a membrane non-permeable to proteins or other macromolecular components, which allows the passage of for example small antimicrobial peptides. The microtiter plate is incubated at 30°C for 2-5 days whereafter anti-fungal activity is analysed by optical density measurements in each well at a suitable wavelength. Low optical density indicates the presence of an anti-fungal activity in the well.

Example 4

Use of very dilute media for HTS screening of enzyme-secreting fungi for improved yield.

In order to mimic the nutrient limiting conditions of fermentor tanks used for production of for example enzymes, microtiter plate-based screens for improved yield are done in dilute growth media. The dilution ensures that the nutrients, rather than the oxygen, become the growth limiting factors. This allows a simplification in the handling of the assay since the dilute substrate can be added directly to the cultures, because the effects of the media on the assay will be very weak.

35 Example 5

Use of very dilute media for HTS screening of enzyme-secreting fungi for improved expression

General HTS set up for yield mutant screening.

5 Mutant libraries of *Aspergillus oryzae* or *Aspergillus niger* strains are screened by HTS to identify mutants having increased expression, i.e. improved yield of an enzyme. Better yield mutants expressing for example peroxidase, laccase, lipase, glucose oxidase, amylase, xylanase, phytase, and 10 aminopeptidase have been identified this way.

For 96-well plate screens, very dilute medium containing 100 - 300 mg/l carbon-source, ensures carbon-limitation without oxygen limitation.

Primary 96-well plate screens involve the dilution of 15 spores from distinct mutant pools into fermentation substrate so that one spore in average is inoculated per well when 100 μ l of medium is dispensed into each well. After inoculation, the plates are incubated for 3-4 days at 30-34°C under static conditions in a culture box (see example 1). The wells are then 20 assayed for enzyme activity, for example by the addition of a substrate solution directly to the growth wells.

Mutants of interest are isolated and single colonies transferred to agar slants. Spores from agar slants are inoculated into 96-well plates with approx. 10³ spores per well and 25 fermented under static conditions for 4 days at 34°C to retest isolates. Subsequently, selected isolates are tested in shake flask fermentations.

Isolation of mutants having increased Shearzyme® expression

30 An *Aspergillus oryzae* transformant MT2181 expressing recombinant Shearzyme®, was mutated by using UV-irradiation. The MT2181 mutant library was fermented in SH-CDY substrate (pH 6.5) in 96 wells plates as described above. The SH-CDY substrate contained pr litre: 0.3 g MgSO₄.7H₂O, 0.3 g K₂SO₄, 5.0 g 35 KH₂PO₄, 0.1 ml tracemetal solution, 0.013 g urea, 0.010 g yeast extract, and 0.1 g maltose. The tracemetal solution contains pr

litre: 13.8 g FeSO₄, 7H₂O, 8.5 g MnSO₄, H₂O, 14.3 g ZnSO₄, 7 H₂O, CuSO₄, 5 H₂O, NiCl₂, 6 H₂O, 3 g citric acid.

Cultures were then assayed for increased yield, i.e. increased Shearzyme® activity. Shearzyme®, which is an endoxylanase, could be quantitated by using a microtiter plate assay and p-nitrophenyl-β-D-xylopyranoside (Sigma N-2132) as substrate, since the enzyme possesses some exo activity, and yellow p-nitrophenol was generated. A 100 mg/ml p-nitrophenyl-β-D-xylopyranoside stock solution was prepared in DMSO. Subsequently, working substrate was prepared by diluting the stock 10 folds in 0.1 M phosphate buffer (pH 6.0). Standard Shearzyme® (BioFeed® (S), Novo Nordisk A/S, Bagsværd, Denmark) was prepared so as to contain 20 FXU/ml in 0.1 M phosphate buffer, pH 6.0. The standard was stored at -20°C until use. Shearzyme® stock was diluted appropriately to obtain a standard series ranging from 0.1 to 5.0 FXU/ml just before use. 25 µl samples were dispensed to wells in 96-well plates followed by 120 µl of substrate by using a robotic set up. Using a plate reader, the absorbance at 405 nm was recorded as the difference of two readings taken at approximately 4 minutes intervals. Shearzyme® units/ml (FXU/ml) were calculated relative to the Shearzyme® standards.

In the 96-well primary screen followed by the re-test in 96-well plates, 9 mutants from the "MT2181 mutant library" were selected. These mutants produced higher yields of Shearzyme® than the control strain *Aspergillus oryzae* MT2181. The 4 Shearzyme® producing mutants with highest yields were evaluated in shake flasks. The results obtained are shown in table 2 below where the Shearzyme® yield of *A. oryzae* MT2181 as a control is normalized to 100.

Strain #	96-well re-test (FXU/ml)	Shake flask (FXU/ml)
MT2181-3-14	148	118
MT2181-5-16	140	121
MT2181-6-01	126	130
MT2181-7-23	130	125

Table 2. Shearzyme® expression by MT2181 Mutants

As shown in table 2, the 4 selected mutants produce approximately 20-30 % more Shearzyme® than the control strain A. 5 oryzae MT2181, when fermented in shake flasks.

Example 6

Screening for cellulase activity using Dichlorotriazino-5-aminofluorescein (DTAF) labelled cotton swatches.

10

Preparation of Cotton swatches

Style 400 cotton (Center for Test Materials, PO Box 120, 3130 AC Vlissingen, Holland) was cut to appropriate sizes. The cotton swatches were swollen in a 0.1 N NaOH solution for 24 15 hrs. prior to the labelling in order to enhance the accessibility of the hydroxylic groups to the probe. DTAF dissolved in 0.1 N NaOH was added to the swollen swatches, and the reaction mixture was allowed to react for 24 hrs. at room temperature in the dark on a celloshaker-board. The labelled swatches were 20 washed several times using 0.1 N NaOH, water, MeOH, and finally water again. After the last wash-cycle no probe could be found in the supernatant. The swatches were finally dried. The labelled cotton swatches have a clear yellow colour. Further the DTAF-probe makes the swatches highly fluorescent at the 485/518 25 (excitation/emission) wavelength bands.

Similar fluorescence measurements were obtained whether cotton was labelled in swatches of 5-11 g or whether it had been cut into swatches of approx. 3 mg, suitable for subsequent microtiter plate analysis.

Specific release of fluorescence by incubation of DTAF labelled cotton swatches with purified cellulases

Labelled swatches were distributed in each well of a 96-well microtiter plate, washed in deionized water, and incubated with 50 µl cellulase enzyme solution and 150 µl buffer for 2h, at 40°C, 400 rpm. The microtiter plates were centrifuged at 2800 rpm and fluorescence intensity (485/518 nm) was measured on 150 µl of the supernatant, using a spectrophotometer. The cellulases used were endoglucanase (EG) V, EGV-core, EG I, and CBH II from *Humicola insolens* (prepared as described in M. Schülein, J. Biotechnology, vol 57, pp. 71-81, 1997).

Table 3 shows that EG V performs much better per endocellulase activity unit than EG I, EG V-Core or CBH II. This is consistent with the performance of these enzymes in multi-wash experiments to determine their effect. Hence, DTAF-swatches can be used to assay cellulases and the results correlate with the performance of the cellulase in performing a full-scale wash experiments.

Further, the signal-to-noise ratio is improved by pre-washing the cotton twenty times in a laundry machine, using either water or a model detergent, before labelling. Also, the signal-to-noise ratio is improved by increasing the number of labelling cycles.

25

cellulase units/ml	CBH II	EG I	EG V	EG V core
0,0	24,8	29,0	32,6	29,6
0,9	27,2	33,6	37,7	29,7
2,8	24,5	39,2	37,0	29,7
8,0	21,6	39,4	49,3	32,8
25,0	23,8	50,7	70,2	36,4
75,0	36,9	57,6	112,8	61,9

Table 3. Performance of 4 cellulases in the swatch assay expressed as fluorescence intensity in arbitrary units (example 6).

5 High throughput screening for cellulase activities in cell culture supernatants

Recombinant *H.insolens* cellulase VI (WO9901544) was expressed in yeast cells (*S.cerevisiae*, YNG318) using standard methodologies (Sambrook et al.: "Molecular Cloning. A Laboratory 10 Manual", Cold Spring Harbor, NY, 1989). Transformed cells were grown in SC-ura⁻ media for 3 days at 30°C. 100µL culture supernatant was transferred to an assay plate containing labelled cellulose swatches as described above. After an appropriate time of incubation the well supernatants were transferred to an 15 assay plate, and the fluorescence was measured. A control experiment was performed in 0,2M tris buffer at pH 7 employing 100 µl supernatant from yeast cultures expressing either recombinant active cellulase EGVI, or an inactive variant of EGVI. Clones expressing active cellulases gave rise to higher levels 20 of fluorescence than clones expressing the inactive variant:

EGVI(active): 47.8 +/- 1.1 (n=6)

EGVI(inactive): 77.7 +/- 1.9 (n=6)

25 The DTAF labelled cotton swatches can also be used in a cell culture-based screen for active cellulases in a high-throughput format: Yeast cells expressing cellulases are suspended in minimal media, and aliquoted into 96 well plates at an average of about 30 wells with growth per 96 well plate (to 30 obtain predominantly single cell isolates). After 3 days of growth at 30°C, 100µL culture supernatant is transferred to an assay plate containing cellulose swatches as described above. Optionally, a well insert (e.g., Nunc TC Insert) may be inserted, to separate yeast cells and the cellulose swatch. The 35 insert should contain a membrane permeable to the cellulase en-

zyme but not to the cells. After an appropriate time of incubation, the well supernatants (without any prior centrifugation) are transferred to an empty assay plate, and the fluorescence is measured for each well. Clones expressing active cellulases give rise to higher levels of fluorescence.

Example 7

His-tagging/purification of proteases to achieve equal protein input from differentially expressing clones.

The DNA sequence encoding the protease Savinase[®] (Novo Nordisk A/S, Denmark) is translationally fused to a sequence encoding a His6 tag and libraries of Savinase[®]-His6 variants are produced and introduced into *Bacillus*. After standard growth, a limited number of Savinase[®] enzymes of each variant (about 10% of what is secreted by *Bacillus* carrying the wildtype Savinase gene) are immobilized in the wells of Ni-NTA microtiter plates. The unbound fraction including cells and excess Savinase[®] is removed, the plate washed once or twice in a buffer containing 5-20 mM Imidazole. The His-tagged Savinase[®] variants are released from the solid support by the addition of 250 mM Imidazole, and aliquots of the supernatants from each well are used as input in a wash performance assay as described in the previous examples.

Example 8

Determination of Savinase[®] concentration by fluorescence polarization measurements.

The CI-2 protease inhibitor is labelled by standard means with a fluorescent probe. After growth as described above, the amount of a Savinase[®] variant in a given well may be measured directly in the wells by addition of fluorescence-labelled inhibitor which upon binding to the protease changes rotational speed. The rotational speed may be monitored by fluorescence polarization analysis or by other means which monitors diffusion (e.g. fluorescence correlation spectroscopy). Since the amount of Savinase[®] variant may vary significantly between the

individual wells, the fluorescently labelled CI-2 inhibitor is added in two or three steps of defined amounts and the fluorescence polarisation is measured after each CI-2 addition. The determined concentration of Savinase® variant in the individual well may be used to adjust the input volume from this well into the activity assay, or can be used to correct the obtained activity data, in order to determine the specific activity of the Savinase® variant of that well.

10 **Example 9**

High throughput screening in filamentous fungi

It is a difficult task to make protein libraries in filamentous fungi. One problem is the different morphology and growth rate of different clones. This gives rise to large and 15 small clones, as well as clones that might be outcompeted by other clones. Assaying fungal transformants of different sizes is also a difficult task due to differences in amounts of expressed enzymes. Another problem is the handling of all the clones on the plate.

20 These problems can be overcome by growing the individual clones in alginate balls with a high molecular weight (HMW) dextran as carbon source, and then using the balls as the source of clonal cell cultures in a microtiter plate. The HMW dextran lets the slow-growing transformants outgrow, without cross 25 contamination from faster growing transformants.

A Polymerase chain reaction, using pAHL (carrying a Lipase gene) as the template and 2 pmol/ml of each primer: oligo 115120: (GCTTTGTGCAGGGTAAATC), and oligo 134532: (GAGCAATAT-CAGGCCGCGCACG) is run under the following conditions: 94°C, 5 30 min.; 30 cycles of (94°C, 30 sec.; 50°C, 30 sec.; 72°C, 1 min.), and 72°C, 5 min. and a commercial Taq polymerase, such as AmpliTaq, (Perkin-Elmer Corp., Branchburg NJ, USA). For example, the PCR conditions may result in a high rate of error, and therefore a library of Lipase mutants are generated.

35 Protoplasts of the filamentous fungi strain JaL250 is transformed using 2µg of pENII1298, which had been digested with

Ball and SgrA1 to remove most of the lipase encoding sequence, and 5 μ g of the above PCR product. The vector and the PCR fragment are allowed to recombine *in vivo* as described in WO 97/07205.

5 The protoplasts are washed twice with ST (0.6M sorbitol, 100mM TRIS pH7.0) in order to remove CaCl₂. The protoplasts are then re-suspended in an alginate solution (1.5% alginate, 2 % high molecular weight dextran (5-40*10⁶ kd), 1.2 M sorbitol, 10 mM TRIS pH7.5). Using a pump, the suspension is pumped through
10 a small nozzle where small suspension droplets are made. These drops fall down (15 cm) into a shake flask (500ml) containing 0.2M CaCl₂, 1.2 M sorbitol, 10 mM TRIS pH7.5. Droplets of an alginate solution (typically 1-2 %) turn into hard balls when they encounter a CaCl₂ solution (such as a 0.2 M solution).
15 Small alginate balls of the size 2.5 mm in diameter are generated by this procedure. The protoplast suspension should be made so that approximately 1 out of 5 balls contains a transformed protoplast, in order to avoid multiple clones in the same ball.

20 The protoplast-containing alginate balls are grown overnight at 30°C degrees in STC (10 mM CaCl₂, 1.2 M sorbitol, 10 mM TRIS pH7.5) in order to regenerate the cell wall. After a couple of washes in sterile water to remove sorbitol, the balls are transferred to 1* Vogel media and grown 2-3 days at 30°C degrees. The sole carbon source is the dextran in the balls. This prevents cross-contamination from ball to ball, and allows slow-growing transformants to gain reasonable biomass.

The alginate balls are transferred to a microtiter plate; one ball into each microtiter well. Liquid growth media is
30 added to the wells and after a period of growth incubation, the growth media is analyzed for lipase activity under the desired conditions.

Claims

1. A method for high throughput screening (HTS) of a large population of host cells for production of a molecule of interest, the method comprising the steps of:

- 5 a) arranging the host cells in a spatial array so each position in the spatial array is occupied by one cell,
- b) cultivating the host cells under growth conditions suitable for HTS,
- c) assaying each array position for production of the molecule of interest, and
- 10 d) selecting the cells from those positions where the molecule was produced, as determined in step c).

2. The method according to claim 1, wherein the spatial array
15 of point a) is a microtiter plate, a solid surface or a textile surface.

3. The method according to claim 1 or 2, wherein after step b)
a sample is transferred from each position in the spatial array
20 to a position in a second spatial array which is then used onwards in the method, preferably the second array is a microtiter plate, more preferably a solid surface, and most preferably a textile surface.

25 4. The method according to any of claims 1 - 3, where each position in the spatial array is occupied by a bead comprising one cell, preferably the bead is an agarose-bead.

5. The method according to any of claims 1 - 4, wherein the
30 molecule of interest in claim 1 step c) is assayed in either whole broth, supernatant of cells that secrete the molecule, a lysate of cells that produce the molecule, or is assayed while still inside cells that produce the molecule.

6. The method according to any of claims 1 - 5, wherein the host cells are of mammalian origin, preferably hybridoma cells.

7. The method according to any of claims 1 - 5, wherein the host cells are of microbial origin, preferably bacterial or fungal.

8. The method according to claim 7, wherein the host cells are chosen from the group consisting of *Escherichia coli*, *Bacillus subtilis*, *Bacillus licheniformis*, *Bacillus clausii*, *Aspergillus niger*, *Aspergillus oryzae*, *Aspergillus nidulans*, and *Saccharomyces cerevisiae*.

9. The method according to any of claims 1 to 8, wherein the molecule of interest is chosen from the group consisting of polypeptides, small peptides, lipopeptides, antimicrobials, small molecules and pharmaceutical molecules.

10. The method according to any of claims 1 to 8, wherein the molecule is an enzyme, preferably chosen from the group consisting of amylase, amyloglucosidase, carbohydrase, carboxypeptidase, cellulase, glucoamylase, glucose oxidase, glucosidase, haloperoxidase, hemicellulase, invertase, isomerase, laccase, lipase, lyase, mannosidase, oxidase, pectinolytic enzyme, peroxidase, phytase, phenoloxidase, polyphenoloxidase, protease, and transglutaminase.

11. The method according to any of claims 1 to 10, wherein the growth conditions of claim 1 step b) ensure minimal variation in the molecule concentration between all positions in the spatial array.

12. The method according to any of claims 1 to 11, wherein the assay of claim 1 step c) is done under conditions where the assay is essentially independent of the concentration of the molecule.

13. The method according to claim 12, where the molecule is an enzyme, and where the enzyme concentration is in a range of preferably 98% below to 400% above the average concentration of enzyme employed in the assay, more preferably 50% below to 200 % above the average concentration of enzyme employed in the assay.

14. The method according to any of claims 1 to 13, wherein the assay done in claim 1 step c) is followed by the additional steps of:

- a) determining the concentration of the molecule of interest for each position; and
- b) adjusting the concentration of molecules in each position to essentially the same level and then performing the assay of claim 1 step c); or
- c) performing the assay in claim 1 step c) and then correcting the data obtained with regard to the concentration of the molecule.

20

15. The method according to claim 14, wherein the molecule of interest is an enzyme, and wherein the concentration in claim 14 step a) is indirectly determined by a method chosen from the group consisting of active site titration using an enzyme inhibitor, measurement of fluorescence polarization using a fluorescently labelled enzyme inhibitor, measurement of labelled anti-enzyme antibodies, and measurement of a marker protein co-expressed with the enzyme.

30 16. The method according to any of claims 1 - 15, wherein the cultivation of claim 1 step b) is followed by an additional step of recovering from each position essentially the same amount of molecule and using these recovered amounts in the assay of claim 1 step c).

35

17. The method according to claim 16, wherein the molecule is a polypeptide, and where the recovery is achieved by fusing a polyhistidine-tag to the polypeptide and recovering this construct on a Ni-NTA solid support.

5

18. The method according to any of claims 1 - 17, where the assay used in claim 1 step c) allows the determination of a chemical or physical event in a live cell, preferably lysis, growth inhibition, growth promotion, production of a protein or 10 another molecule by the cell.

19. The method according to any of claims 1 - 17, where the assay used in claim 1 step c) allows the determination of an event in which the molecule of interest interacts with another 15 molecule, preferably leading to a binding event or a chemical modification of the other molecule.

20. The method according to any of claims 1 - 17, where the assay used in claim 1 step c) allows the determination of the 20 cleavage, isomerization, ligation or other modification of a substrate.

21. The method according to claim 20, where the substrate is a polymer, preferably chosen from the group consisting of poly- 25 peptides, celluloses, polysaccharides, and starches.

22. The method according to claim 20 or 21, where the substrate is immobilized to the surface of a solid material, preferably textile or ceramics.

30

23. The method according to any of claims 20 - 22, where the substrate is a solid surface, preferably a cellulose swatch.

24. The method according to any of claims 20 to 23, where the 35 substrate is labelled with a detectable probe.

25. The method according to claim 24, where the probe is fluorescent, preferably fluorescein-5-isothiocyanate (FITC) or di-chlorotriazino-5-aminofluorescein (DTAF).

5 26. The method according to any of claims 1 - 25, where the assay of claim 1 step c) is performed in a solution comprising a detergent and where the molecule of interest is an enzyme.

10 27. The method according to any of claims 1 - 26, where the assay used in claim 1 step c) involves detection of a signal chosen from the group consisting of fluorescence, fluorescence polarization, absorbance, radioactivity, nuclear magnetic resonance (NMR) and luminiscence.

15 28. The method according to any of claims 1 - 27, where the growth conditions of claim 1 step b), have been optimized to give minimal variation in the amount produced of the molecule of interest in each position of the spatial array.

20 29. The method according to any of claims 1 - 28, where the growth conditions are chosen so as to mimic the conditions under full scale fermentations, preferably highly dilute media are chosen.

25 30. A method for screening a DNA library for DNA of interest, the method comprising the steps of:

- a) creating host cells comprising the DNA library,
- b) using the method described in claim 1 to assay for a host cell producing a molecule of interest; wherein the host cell comprises the DNA of interest.

30 31. The method according to claim 30, wherein the DNA library is generated from a natural DNA sequence by DNA shuffling or directed evolution.

32. The method according to claim 30, wherein the host cells of step a) are of microbial origin.

33. A recombinant vector comprising DNA isolated by a method as defined in any of claims 30-32.

34. A recombinant host cell comprising DNA isolated by a method as described in any of claims 30-32 or the vector according to claim 33.

10

35. The cell according to claim 34, which is of microbial origin.

36. The cell according to claim 35, which is of *Bacillus* or *Aspergillus* origin.

37. The cell according to claim 36, which is chosen from the group consisting of: *Escherichia coli*, *Bacillus subtilis*, *Bacillus licheniformis*, *Bacillus clausii*, *Aspergillus niger*, *Aspergillus oryzae*, *Aspergillus nidulans*, and *Saccharomyces cerevisiae*.

38. A transgenic animal containing and expressing the DNA of interest according to any of claims 30-32.

25

39. A transgenic plant containing and expressing the DNA of interest according to any of claims 30-32.

40. A method of producing a molecule of interest, which method comprises cultivating a cell according to any of claims 34-37 in suitable culture medium under conditions permitting expression of the DNA of interest and recovering the resulting molecule from the culture medium.

68 .

41. A method of producing a material of interest, which method comprises recovering the molecule from any part or secrete of/from the transgenic animal according to claim 38.

5 42. A method of producing a molecule of interest, which method comprises growing a cell of a transgenic plant according to claim 39, and recovering the molecule from the resulting plant.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/DK 00/00568

A. CLASSIFICATION OF SUBJECT MATTER

IPC7: C12N 9/00, C12M 1/18

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC7: C12Q, C12M, C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

SE,DK,FI,NO classes as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 9749987 A1 (CELLSTAT TECHNOLOGIES, INC.), 31 December 1997 (31.12.97), page 6, line 5 - line 16 --	1-42
X	WO 9955826 A1 (GENOVA PHARMACEUTICALS CORPORATION), 4 November 1999 (04.11.99), page 7, line 8 - line 18 --	1-42
A	WO 9640974 A1 (PRESIDENT FELLOWS OF HARVARD COLLEGE), 19 December 1996 (19.12.96), abstract --	1-42

 Further documents are listed in the continuation of Box C. See patent family annex.

- * Special categories of cited documents
- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier application or patent but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed
- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- "&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
5 March 2001	07.03.2001
Name and mailing address of the ISA/ Swedish Patent Office Box 5055, S-102 42 STOCKHOLM Facsimile No. + 46 8 666 02 86	Authorized officer Frida Plym Forshell/mj Telephone No. + 46 8 782 25 00

INTERNATIONAL SEARCH REPORT

International application No.

PCT/DK 00/00568

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 9960170 A1 (HYSEQ, INC.), 25 November 1999 (25.11.99), page 6, line 27 - page 8, line 15 -- -----	1-42

INTERNATIONAL SEARCH REPORT

International application No.
PCT/DK00/00568

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:

2. Claims Nos.: **11, 12, 28** because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
see next page

3. Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- The additional search fees were accompanied by the applicant's protest.
 No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/DK00/00568

Present claims 11, 12 and 28 relate to a method for high throughput screening defined by reference to desirable characteristics, namely growth conditions that ensure minimal variation in the molecule concentration between different positions in an array and an assay that is independent of the molecule concentrations. The claims cover all methods having these characteristics whereas the application provides support within the meaning of Article 6 PCT and / or disclosure within the meaning of Article 5 PCT for only a very limited number of such methods. Thus, a meaningful search over the whole of the claimed scope is impossible.

Consequently, the search has been carried out for those parts of the claims that appear to be supported and disclosed, namely the methods described in examples 1, 2 and 9 that define how to reach the desired characteristics.

INTERNATIONAL SEARCH REPORT

Information on patent family members

05/02/01

International application No.

PCT/DK 00/00568

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO 9749987 A1	31/12/97	AU CA EP	3508197 A 2258279 A 0950184 A	14/01/98 31/12/97 20/10/99
WO 9955826 A1	04/11/99	AU	3753599 A	16/11/99
WO 9640974 A1	19/12/96	AU US	5872096 A 5654150 A	30/12/96 05/08/97
WO 9960170 A1	25/11/99	AU	4194899 A	06/12/99